UNIVERSIDAD DE CHILE.
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS.
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL.

PLATAFORMA DE ANÁLISIS E IDENTIFICACIÓN DE DEMANDA DE COMPETENCIAS LABORALES EN LOS AVISOS DE TRABAJO DE LA BOLSA NACIONAL DE EMPLEO MEDIANTE TÉCNICAS DE TEXT MINING

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERIO CIVIL INDUSTRIAL

JAVIER IGNACIO MOLINA SALGADO

PROFESOR GUÍA:

SR. JUAN D. VELÁSQUEZ SILVA

MIEMBROS DE LA COMISIÓN:

SR. BENJAMÍN VILLENA ROLDAN

SR. FELIPE VILDOSO CASTILLO

Este trabajo ha sido financiado parcialmente por Proyecto FONDECYT Regular 1151479 y Proyecto CONICYT PIA SOC 1402.

SANTIAGO DE CHILE

2018

Resumen Ejecutivo

RESUMEN DE LA MEMORIA
PARA OPTAR AL TITULO DE:
Ingeniero Civil Industrial.
POR: Javier Ignacio Molina Salgado.
FECHA: 17/01/2018
PROF. GUÍA: Juan Velásquez Silva.

La Bolsa Nacional de Empleo (BNE), es una plataforma laboral estatal, pública y gratuita. Esta plataforma sirve de intermediario entre empleadores y personas que se encuentran en búsqueda de trabajo. Esta iniciativa depende del Ministerio del Trabajo y Previsión Social en conjunto con el SENCE, Servicio Nacional de Capacitación y Empleo.

Por otro lado, existen competencias laborales, las cuales son definidas como habilidades, aptitudes y conocimientos que necesarios para la ejecución de alguna función dentro de una actividad laboral o puesto de trabajo. Estas son requeridas por los trabajos para ser correctamente ejecutados y también son poseídas por cada trabajador producto de su formación y experiencia.

Esta memoria tiene dos objetivos principales: el primero es detectar las competencias laborales que requieren los avisos de trabajo de la BNE y en segundo lugar diseñar e implementar una plataforma de análisis para la información generada, de manera de disponer de una herramienta de apoyo a la gestión y decisión para las entidades involucradas.

Para cumplir con dichas metas se desarrolló un sistema que utilizó técnicas de text mining, además de integrar un catálogo de competencias laborales junto a un motor de búsqueda para generar coincidencias entre las competencias laborales y los avisos de trabajo. Luego con la información generada se diseñó e implementó una plataforma de análisis (un data warehouse) accesible desde la web y que permite realizar distintos análisis de las competencias laborales demandadas segmentando por factores como temporalidad, sector económico, tipo de trabajo, entre otros.

Finalmente, se desarrolló una herramienta de detección de competencias laborales y una plataforma para su análisis a niveles agregados o segmentados con una precisión del 95% para los grupos más altos de coincidencias.

Las conclusiones de esta memoria apuntan a utilizar el trabajo desarrollado como base para nuevas líneas de investigación y aplicaciones prácticas en las distintas plataformas laborales, así como posibles mejoras al sistema actual.

Agradecimientos

Quiero agradecer a mi familia, a mis padres Oscar y Pilar, quienes me han dado una buena educación y herramientas fundamentales para enfrentar distintos desafíos a lo largo de mi vida, incluyendo el transcurso de mi carrera universitaria.

A mi hermano Oscar por estar conmigo y ser un compañero en muchas situaciones que nos han tocado, generando siempre un ambiente grato para todos.

A mis amigos y compañeros, quienes forman parte importante de mi vida con quienes he pasado grandes momentos en distintas etapas. Con quienes nos permitimos reír y conversar siempre. En especial a José Tomás Lagos, quien ha sido mi compañero desde primero básico hasta el final de nuestras carreras, con quien he estudiado y trabajado codo a codo en la mayoría de los cursos y encontrando siempre espacios para pasarlo bien y reírnos.

A Catalina Barker, por ser el pilar de mi vida, apoyarme siempre sin duda alguna, ser el amor de mi vida y mi mejor amiga. Sin ti nada de esto habría sido posible y es la primera concreción de nuestro futuro.

Al WIC que me ha acogido y brindado distintas oportunidades para realizar diferentes proyectos. A todos sus integrantes que siempre tienen una gran disponibilidad para ayudar y guiar en el desarrollo de estos.

A mis profesores, Juan Velásquez, por la oportunidad y confianza para desarrollar mi memoria, clases auxiliares y otros proyectos. A Felipe Vildoso por ser un gran apoyo en el desarrollo de esta memoria, brindarme oportunidades constantes para embarcarse en nuevos emprendimientos, ayudarme cada vez que lo necesite y por su amistad.

Finalmente, a todos quienes me han sacado una sonrisa en el transcurso de mi carrera y de quienes he podido aprender muchas cosas y distintas visiones de la vida. No puedo olvidar agradecer a Milo y Almendra por hacer mi vida más feliz, ser fieles y brindarme su compañía siempre.

Tabla de Contenido

Índice de Tablas

Índice de Ilustraciones

Capítulo 1

Introducción

1.1. Antecedentes generales.

El mercado laboral es un mercado friccional, donde la fuerza laboral y las empresas buscan en el otro al candidato adecuado para los puestos de trabajo, así como los empleadores buscan empleos que se adecuen a sus necesidades y preferencias. Por su naturaleza se espera que no exista un calce perfecto entre las competencias laborales ofrecidas y las requeridas en el mercado.

A lo anterior se debe agregar que en las últimas décadas el mercado laboral a nivel mundial se ha visto afectado por distintos factores como; globalización económica, penetración de nuevas tecnologías, aumento de la migración, entre otros. Estas situaciones han causado que las competencias laborales demandadas en el mercado laboral también varíen rápidamente. Dichas variaciones tienden a aumentar la brecha entre las competencias laborales demandas en el mercado laboral (de parte los empleadores en general), las desarrolladas o disponibles en las nuevas generaciones de trabajadores y aquellas que han ido quedando obsoletas en el tiempo. [17] [3]

La brecha anteriormente mencionada tiene un impacto a través de varios costos al mercado y a la productividad de las empresas. [18]

Por ejemplo, en el mercado estadounidense esta brecha significa que en promedio hay una demora de 3 meses para contratar un ingeniero, investigador o científico, y más de 2 meses en llenar una vacante para trabajadores capacitados en producción. Se estima que en una década la industria manufacturera estadounidense tendrá cerca de 2.000.000 de vacantes sin satisfacer, esta cifra en 2011 alcanzaba apenas los 600.000. [19]

Otra evidencia del efecto de la brecha entre competencias laborales requeridas y ofertadas en el mercado laboral es el fenómeno llamado efecto histéresis, observado en Europa, el cual consiste en que las competencias laborales de los trabajadores, que perdían sus trabajos durante la crisis europea, quedaban obsoletas en el mercado al momento de su reinserción laboral. Esto generó un impacto aumentando el desempleo estructural, es decir, el desempleo causado por la brecha entre los empleos ofertados y las habilidades existentes en los postulantes. Este desempleo aumento desde un valor cercano y menor al 9% a un 10% en 6 años. [20]

Además, existe un impacto a nivel individual de aquellas personas que no logran hacer un buen emparejamiento de las competencias laborales que poseen con las requeridas por sus puestos de trabajo. Esos trabajadores reciben un menor sueldo en comparación a quienes si logran emparejar sus competencias laborales con su trabajo de mejor forma. Esta diferencia de salario se acentúa en el tiempo de actividad laboral del trabajador a lo largo de sus años de experiencia. [21]

Dado lo expuesto anteriormente, se puede ver que existe un interés y necesidad de trabajar e investigar en distintas herramientas, metodologías y estrategias a niveles locales y de colaboración internacional que impulsen la creación sostenida de empleos y para lograrlo es necesario disminuir las brechas de competencias laborales existentes y prepararse para las futuras demandas de nuevas competencias.

Un aspecto importante de esta tarea es lograr detectar las competencias laborales demandas en el mercado.

1.2. Antecedentes a nivel nacional.

En Chile, a nivel gubernamental, la institución encargada del mercado laboral es el Ministerio del Trabajo y Previsión Social, MINTRAB, que declara su misión:

"Contribuir al desarrollo del país, impulsando políticas públicas que promuevan el trabajo decente, la formación para el trabajo, la seguridad y salud laboral, una mayor integración de grupos vulnerables en el mercado del trabajo, así como los cambios normativos necesarios para la ampliación y ejercicio de los derechos de los trabajadores, especialmente los derechos colectivos. De igual manera, el Ministerio promoverá los cambios necesarios al sistema de previsional."

De ella se destacan para este trabajo de título la labor formativa y el propósito de una mayor integración laboral por parte del ministerio. [26]

En este sentido el ministerio del trabajo se relaciona con el organismo técnico descentralizado SENCE, Servicio Nacional de Capacitación y Empleo. Esta institución declara su misión como:

"Mejorar la empleabilidad de los trabajadores ocupados, personas desocupadas e inactivas, con especial foco en las más vulnerables, a lo largo de su vida laboral, a través de una gestión articulada y con calidad de la orientación, capacitación e intermediación laboral, para contribuir a la productividad de Chile y sus regiones."

El SENCE, dentro de sus funciones incluye la promoción y estimulación de programas de capacitación para los trabajadores del país, además de coordinar las Oficinas Municipales de Información Laboral y fomentar la calidad de las instituciones intermediaras (OTIC) y ejecutoras de las capacitaciones (OTEC).

Esta institución fue creada en el año 1976 como resultado de la promulgación de la Ley Nº 1446, Estatuto de Capacitación y Empleo, y tiene presencia nacional en las 15 capitales

regionales del país. En la Ley de Presupuestos del año 2017 se destinaron $ 240.445 millones de pesos chilenos en becas, bonos de capacitación para micro y pequeños empresarios, subsidios al Empleo Ley Nº 20.338 y subsidio al empleo a la Mujer Ley Nº 20.595. [22] [23]

Según el anuario estadístico público del SENCE correspondiente al año 2016 [24] y tal como se ilustra en la tabla 1.1 se brindaron casi 1.5 millones de capacitaciones en los programas ofrecidos o impulsados por el SENCE. Esto significó un gasto total para el Estado de aproximadamente $ 210.000 millones de pesos chilenos en estas capacitaciones, considerando el gasto privado incurrido de casi $ 34.000 millones de pesos, da un total de aproximadamente $ 245.000 millones de pesos invertido en capacitaciones a través del SENCE con participación de actores privados.

Programa	Participantes[1]	Gasto Publico (Millones)	Gasto Privado (Millones)	Gasto Total (Millones)
Impulsa Personas	1.355.430	$106.423	$33.791	$140.215
+ Capaz	96.925	$92.390	$ -	$92.390
Oficios Registro Especial	5.196	$11.268	$ -	$11.268
Total	1.457.551	$210.082	$33.791	$243.874

Tabla 1.1: Participantes, gasto público y gasto privado del SENCE el año 2016.
Fuente: Elaboración propia, a partir de los anuarios públicos del SENCE.

El SENCE, administra la llamada Franquicia Tributaria de Capacitación, que es un incentivo tributario a las empresas que les permite descontar de sus impuestos hasta un 1% de la planilla anual de remuneraciones imponibles los gastos relacionados con capacitaciones, evaluaciones y certificación de competencias laborales de sus empleados. Los cursos de capacitaciones que apliquen a la Franquicia Tributaria de Capacitación deben ser aprobadas previamente por el SENCE. [44]

Además, el SENCE para poder guiar, justificar y estandarizar sus programas formativos necesita definir varios componentes del mercado laboral que permitan esclarecer qué competencias laborales de las personas son necesarios impulsar y reforzar mediante capacitaciones. De esta forma busca mitigar los efectos y brechas de competencias laborales en el mercado chileno, como los descritos en la sección anterior de Antecedentes Generales.

Entre estos componentes se encuentran las competencias laborales del mercado y perfiles laborales, que corresponden a las competencias laborales necesarias para la ejecución de una actividad laboral particular.

En la definición de los componentes descritos toma parte la Comisión del Sistema Nacional de Certificación de Competencias Laborales, ChileValora. Entre sus tareas está contribuir a que la oferta del sistema público de capacitaciones sea diseñada sobre la base de estándares de competencias laborales que defina el sistema, y según su declaración [25], disminuir la brecha

[1] *La cantidad de participantes considera la cantidad de veces que una persona estuvo inscrita a algún curso de capacitación ese año, no a la cantidad de personas individuales que cursaron en el año 2016.*

de competencias laborales entre los trabajadores y la demanda de las empresas. Se incluye además en sus labores, poner a disposición publica la información de las competencias y perfiles laborales presentes en el sistema, con objetivo de que empresas e instituciones educativas tengan un recurso que les indique lo que necesita el mercado.

El presupuesto público destinado a la labor de ChileValora en el año 2016 fue aproximadamente de $ 2.400 millones de pesos chilenos. [40]

1.3. Motivación

Dado el contexto general y a nivel nacional explicados anteriormente es que surge la necesidad de elaborar este trabajo de título, como una primera aproximación para detectar las competencias laborales en el mercado y servir como una herramienta de gestión a las instituciones públicas chilenas.

Estas instituciones podrían contar con un recurso valioso para sus labores de capacitación, certificación y administración de fondos para sus programas, mediante el apoyo a la labor y metodología actual para la definición de competencias laborales.

1.4. Hipótesis de Investigación.

La hipótesis de investigación planteada consiste en que es posible detectar las competencias laborales requeridas en un aviso de trabajo de la BNE con herramientas de text mining y organizar y presentar la información agregada a través de una plataforma de inteligencia de negocios.

1.5. Objetivos.

Los objetivos generales y específicos para el desarrollo de este trabajo reflejan el propósito del trabajo de título y en segundo lugar los hitos necesarios para lograr dicho objetivo principal.

1.5.1. Objetivo General.

Elaborar una plataforma que mediante text mining y otras herramientas permita la detección de competencias laborales en avisos de trabajos además de la posibilidad de realizar estudios diferenciados por las distintas características propias de los avisos de trabajo, su contexto y las cualidades de las competencias laborales.

1.5.2. Objetivos Específicos.

Para lograr el objetivo general es necesario cumplir con metas o hitos que contribuyan al cumplimiento de dicho objetivo. Estos son presentados a continuación:

- Definir el modelo de datos.
- Definir metodología de text mining a implementar.
- Implementar metodología de text mining en los campos que se requiera.
- Desarrollar el proceso de ETL[2].
- Implementar plataforma data warehouse que permite análisis agregado.
- Validar con los clientes.

1.6. Metodologías.

La metodología utilizada en la realización de este trabajo de título fue la siguiente:

En primer lugar, se realiza un estudio e investigación de la literatura relacionada con la justificación y motivación del trabajo de título, comprendiendo los efectos de la problemática y decantando en un objetivo que aporta herramientas para mitigar dichos efectos de la brecha de competencias laborales requeridas y ofrecidas en el mercado laboral.

En segundo lugar, habiendo definido los objetivos del trabajo de título y una posible solución, se hace un estudio del estado del arte relacionado con las metodologías y áreas posibles para la implementación de la solución y desarrollo del trabajo de título.

Luego se analizan y evalúan las posibles vías de desarrollo del trabajo a realizar y se decide por una metodología para el trabajo venidero.

Después de haber elaborado un sistema de detección de competencias laborales y haberlo implementado en el sistema data warehouse, se procede a validar la solución con los clientes de esta memoria.

Finalmente se concluye sobre el trabajo realizado, se proponen distintas líneas investigativas a partir de lo elaborado y posibles mejoras al trabajo de título. El aspecto principal de esta etapa es la validación o rechazo de la hipótesis investigativa y su fundamento.

---

[2] *De la sigla del inglés Extraction, Transformation y Load. Proceso de extracción, transformación y carga de datos entre sistemas o componentes tecnológicas.*

1.7. Alcances.

Se debe considerar que el análisis abarca la información provista en la base de datos de la BNE, donde el análisis se focaliza en el área de texto libre del aviso de trabajo, por lo que se puede aplicar en otros repositorios de avisos de trabajos, pero esta memoria no extrapolará el análisis más allá del set de información de la BNE. El repositorio utilizado contempla información entre el 2011 y mediados del 2016, tiempo en el cual la administración deja de ser del sitio privado, www.trabajando.com, y mediante licitación pública pasa a una nueva empresa.

Esta base de datos no podrá ser actualizada dada la arquitectura elaborada, por lo que el análisis es sobre información pasada y no a tiempo real.

Además, dada la naturaleza de la Bolsa Nacional de Empleos hay que tomar en cuenta que el análisis mostrará aspectos sobre el mercado laboral chileno y en particular de los tipos de trabajos que son publicados en la BNE, el sistema podría eventualmente replicarse en otros mercados, pero la metodología para replicar este trabajo considerando los factores de otros mercados y sistemas queda excluida del trabajo de título. Un estudio realizado por Banfi, Choi, Galeguillos, Villena (2017) concluye que los puestos ofrecidos en la BNE son de menor calificación en promedio en comparación a otros portales como trabajando.com. [48]

A pesar de ser un análisis exclusivo a un set local de datos, este trabajo es atractivo para ser implementado en un sistema en línea. Este aspecto no fue considerado en el trabajo de título, en parte, debido a que el sistema de administración del sitio de la BNE ha cambiado su administración desde el momento que entregó la base de datos usada.

Por otro lado, el sistema de información entregable para el análisis agregado de los avisos de trabajo no necesariamente será de uso público, pero sí tendrán acceso ambas partes participes de esta memoria, el centro de investigación WIC y el profesor Benjamín Villena. Además de quedar plasmado en la memoria la metodología y arquitectura utilizada para replicar el sistema.

1.8. Estructura del informe.

A continuación, se presenta y explica la estructura del presente informe:

Un primer lugar se encuentra el capítulo correspondiente a la introducción del trabajo de título, en este capítulo se destallan el contexto general y nacional que dan pie a la motivación y estructuración de este proyecto, se presenta la hipótesis de investigación, objetivos generales y específicos, la metodología a utilizar junto con los alcances de la memoria.

El segundo capítulo detalla el marco teórico en el cual se desenvuelve el trabajo de título, aquí se encuentra una recopilación y explicación de los aspectos básicos para entender el trabajo realizado a lo largo de la memoria. Se detallan conceptos y actores del mercado laboral, metodologías de text mining atingentes y una explicación de un sistema data warehouse y sus componentes.

El tercer capítulo se encarga de explicar, desde una perspectiva general, la estructura tecnológica utilizada y como estas se relacionan e integran para formar el sistema de detección y análisis de demanda de competencias laborales.

El cuarto capítulo tiene la misión de explicar detalladamente el diseño e implementación del sistema encargado de detectar las competencias laborales en el repositorio de avisos de trabajo de la BNE. En este capítulo también se explican los distintos procesos necesarios para configurar e implementar dicho sistema y finalmente se muestra la validación de la información generada en el sistema.

El quinto capítulo corresponde a una explicación detallada del diseño y la implementación de la segunda parte del sistema general, el cual corresponde a la plataforma de análisis de las competencias laborales demandadas. En este capítulo se incluyen las configuraciones necesarias, los modelos utilizados, los procesos para hacer inserción de la información en dichos modelos y, finalmente, cómo realizar consultas a la plataforma.

El sexto capítulo tiene la tarea de reportar los resultados y entregables, producto del desarrollo de la presente memoria, en este capítulo se incluyen los resultados y estadísticas del sistema de detección de competencias laborales, en segundo lugar, se reportan y analizan los resultados de la clasificación hecha para los puntajes asociados a las coincidencias encontradas en el sistema y en tercer lugar los resultados y entregables correspondientes a la plataforma de análisis.

En el séptimo capítulo se explican los distintos aspectos que pueden mejorar el sistema actual desarrollado desde distintos factores, además se muestran propuestas para la ampliación del alcance de la memoria mediante la integración de nuevos repositorios de competencias laborales y avisos de trabajo. Posteriormente, se plantean nuevas líneas de investigación y aplicaciones desde el trabajo desarrollado en esta memoria.

Por último, para finalizar el informe se presenta el octavo capítulo. Este capítulo concluye respecto al cumplimiento de los objetivos generales y muestra recomendaciones propuestas para el trabajo futuro a partir de esta memoria.

Capítulo 2

Marco Teórico

En este capítulo se presentan los conceptos que conforman el marco teórico o conceptual utilizado en el desarrollo de este trabajo, se abarcan definiciones de competencias y aspectos laborales. Luego se presentan definiciones y conceptos relacionados con las técnicas de text mining utilizadas. Finalmente se entregan conceptos del sistema data warehouse implementado.

Todo este marco es necesario para comprender los temas discutidos y técnicas implementadas sin tener una base de conocimiento previa en el tema.

2.1. Mercado laboral.

El mercado laboral es un arreglo institucional en el cual trabajadores y empleadores interactúan, los empleadores buscan la mejor opción para satisfacer sus vacantes y los trabajadores compiten por obtener el mejor puesto de trabajo dada sus preferencias.

Este mercado por naturaleza es friccional, es decir, ambas partes incurren en un gasto de tiempo y recursos para encontrar su contraparte. Además, no hay claridad al respecto de si estas búsquedas generan el mejor resultado para cada empresa o trabajador. Del mismo modo, tampoco es claro que las asignaciones producidas en general sean las socialmente óptimas.

Este trabajo tiene como alcance los avisos publicados en la BNE, durante un tiempo dado, estos representan solo un espacio de interacción entre las partes. Existen otras plataformas laborales, como las presentadas en la Tabla 2.1. Además de otras formas de interacción, como lo son redes sociales, contactos personales, ferias de trabajo y otros tipos de bolsas laborales a nivel de instituciones educativas, agencias privadas, municipalidades, asociaciones profesionales, gremiales entre otras.

A continuación, se presentan y definen distintos conceptos del ámbito de un mercado laboral que serán utilizados a lo largo de esta memoria.

2.1.1. Competencias laborales.

Una competencia laboral es un término que abarca habilidades, aptitudes y conocimientos que son involucrados en las funciones necesarios para llevar a cabo una actividad laboral (en la ilustración 2.1 se representa este concepto). En Chile, a mediados del año 2008, se promulgó la ley para la creación del Sistema Nacional de Certificación de Competencias Laborales, Ley Nº 20.267 [1]. En el artículo 2ª de la ley se define competencia laboral como: "Aptitudes, conocimientos y destrezas necesarias para cumplir exitosamente las actividades que componen una función laboral, según estándares definidos por el sector productivo."

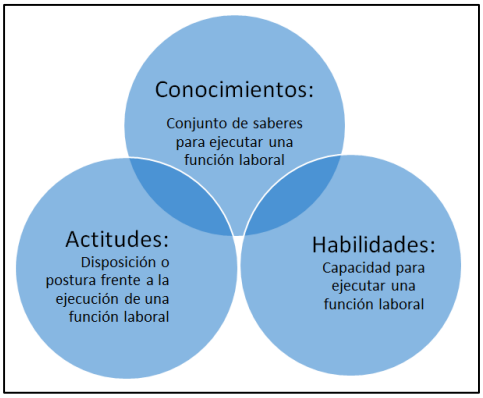


Ilustración 2.1: Diagrama de competencias laborales.
Fuente: ChileValora, Comisión Sistema Nacional Certificación de Competencias Laborales.

La anterior definición coincide también con la propuesta por la institución internacional, ESCO: European Skills/Competences, Qualifications and Ocupations (Clasificación Europea de Capacidades/competencias, cualificaciones y ocupaciones.) [2].

En el marco de la ley chilena anteriormente mencionada se define Unidad de Competencia Laboral [3] (UCL): "Es un estándar que describe los conocimientos, las habilidades y aptitudes que un individuo debe ser capaz de desempeñar y aplicar en distintas situaciones de trabajo, incluyendo las variables, condiciones o criterios para inferir que el desempeño fue efectivamente logrado."

2.1.2. Trabajo u Ocupación.

Un trabajo u ocupación será entendido para este proyecto como un conjunto de competencias laborales o siguiendo la definición de la Ley Nº 20.267 [1] un conjunto de UCL. Este conjunto es requerido por la actividad laboral, trabajo u ocupación para ser ejecutado en los estándares esperados.

[3] *Desde ahora se referirá a una Unidad de Competencia Laboral como UCL.*

Esta definición da paso a definir a un trabajador o postulante a un trabajo.

2.1.3. Trabajador o postulante.

Habiendo entendido que un puesto de trabajo es un conjunto de competencias laborales a forma de requerimiento para ser ejecutado, se da paso a definir a un trabajador o a un postulante a un trabajo como un conjunto de competencias laborales que la persona posee y tiene a disposición para ser utilizados en alguna actividad laboral.

2.1.4. Brecha de competencias laborales.

En el mercado laboral existen puestos de trabajos que deben ser satisfechos por la fuerza laboral, estos trabajos requieren de un conjunto de competencias laborales para ser correctamente ejecutados.

Considerando la suma de todos los conjuntos requeridos por parte de las empresas para cada puesto ofrecido es que se conforma la demanda total de competencias laborales en el mercado.

En cambio, cada trabajador posee un conjunto de competencias laborales a disposición, adquirido a través de su formación y experiencia. La compilación de dichos conjuntos ofrecidos por cada trabajador conforma la oferta total de competencias laborales en el mercado.

Luego se puede suponer que existe información imperfecta en este mercado desde varios frentes. Por un lado, las empresas no pueden conocer perfectamente las competencias laborales de un trabajador antes de ser contratados, tanto la presencia de ellas como su calidad e incluso no existirá un conocimiento perfecto de ellas luego de su contratación.

Por otro lado, cada trabajador no conoce necesariamente sus habilidades o no tiene conciencia de ellas al momento de buscar un trabajo.

Además, no todas las empresas conocen perfectamente que competencias laborales son requeridas para la ejecución de sus puestos de trabajo.

Y como en todo mercado con información imperfecta puede existir una sobre oferta o sobre demanda de competencias laborales. Estas descoordinaciones se denominan brechas de competencias laborales. [3]

2.1.5. Plataformas Laborales.

Una plataforma laboral online es un sitio web que sirve de intermediario entre las compañías que desean encontrar postulantes idóneos para sus puestos de trabajos, mediante la publicación de avisos de trabajos en dichos sitios web y posibles postulantes que buscan empleos acordes a sus intereses y requisitos. Estas datan de principios de los años 90, como documenta Nominet Trust (el financiador de emprendimiento tecnológicos líder en Reino Unido [12]) mediante la

penetración de internet su uso fue incrementando exponencialmente y nuevos actores fueron entrando en el mercado. Ya para el año 2009, cuatro de cada cinco personas que buscaban empleo usaban internet como parte de su búsqueda, concentrándose en personas mayoritariamente jóvenes. [11]

En Chile estas plataformas están consolidadas, hoy existen muchas opciones. Por ejemplo: al buscar empleo se usa cada vez menos la modalidad de dejar los curriculum vitae en formato físico en el lugar donde se está postulando a un cargo, esto de acuerdo a la revista Capital Humano de Emol. [13]

Nombre.	Sitio web.
Trabajando	www.trabajando.com
Laborum	www.laborum.com
FirstJob	www.firstjob.com
Pegas con Sentido	www.pegasconsentido.cl
Empleos Públicos	www.empleospublicos.cl
Chiletrabajos	www.chiletrabajos.cl
Servisenior	www.servisenior.cl
Económicos de El Mercurio	www.economicos.cl/rm/empleos
Bolsa Nacional de Empleo	www.bne.cl

Tabla 2.1: Principales plataformas laborales en Chile.
Fuente: Capital Humano, Emol.

En la tabla 2.1 se presentan los principales sitios web para búsqueda de empleos, estos varían en su enfoque, público objetivo y tipos de empleos, en particular el último elemento, Bolsa Nacional de Empleo, es crucial para el desarrollo de esta memoria.

Bolsa Nacional de Empleo.

La Bolsa Nacional de Empleo, BNE[4], es un sistema informático público y gratuito que presta el servicio de intermediación laboral, tiene como objetivo aumentar la reinserción laboral de los trabajadores cesantes.

Esta iniciativa pertenece al Ministerio del Trabajo y Previsión Social, MINTRAB y del SENCE, Servicio Nacional de Capacitación y Empleo. Su creación se establece en la Ley del Seguro de Cesantía Nº 17.928 en el artículo Nº 61.

Dentro de sus labores, la Bolsa Nacional de Empleo pone a disposición una plataforma laboral. En esta plataforma las empresas y personas jurídicas pueden publicar avisos de trabajos y cualquier persona puede postular a ellos a través del sitio web con una previa inscripción. Su funcionamiento empieza el primer semestre del año 2001 y sigue vigente hasta la actualidad. [14] [45]

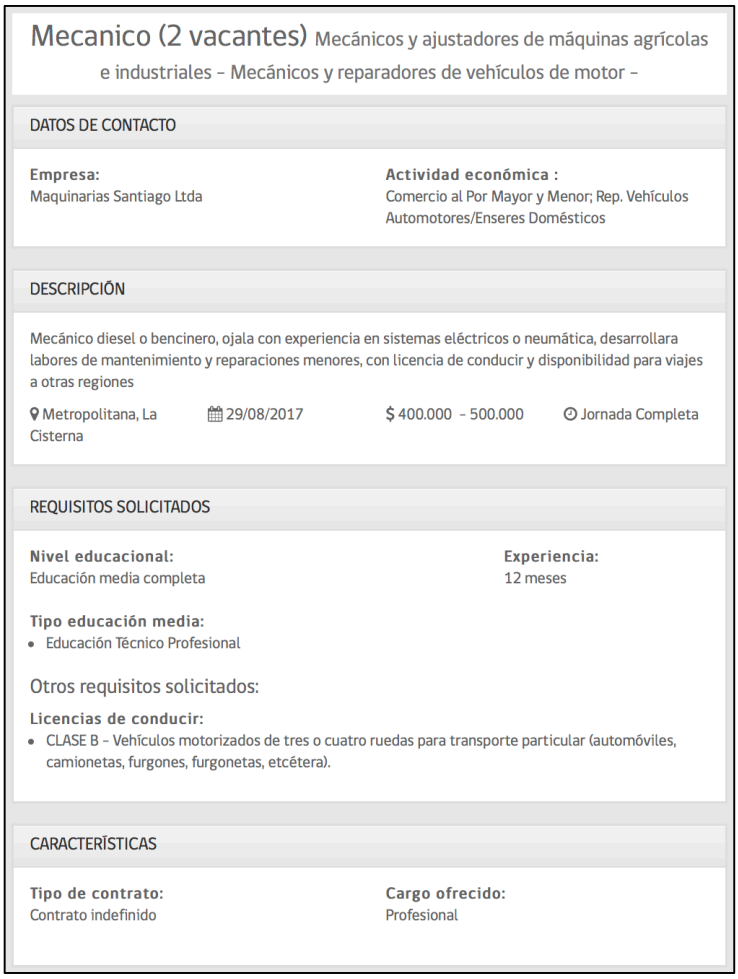
[4] *Desde ahora se refiere a la Bolsa Nacional de Empleo como BNE.*

El uso de esta plataforma genera información sistemáticamente acerca de las ofertas de trabajo publicadas, las interacciones de sus usuarios y sus postulaciones. Mediante un convenio de colaboración suscrito entre el SENCE y la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile, se obtuvo la base de datos de la BNE

Dicha base de datos posee información propia de la plataforma y su funcionamiento, además de información acerca de casi 700.000 avisos de trabajos publicados entre el año 2011 y mediados del 2016.

2.1.6. Aviso de Trabajo.

A continuación, se explica la estructura y contenido de los avisos de trabajos publicados en la BNE. En la ilustración 2.2 se puede observar un ejemplo de aviso de trabajo publicado en la plataforma de la BNE, en el cual se especifican distintos requerimientos e indicaciones para informar a los postulantes acerca del empleo que se está ofreciendo.


Mecanico (2 vacantes) Mecánicos y ajustadores de máquinas agrícolas e industriales – Mecánicos y reparadores de vehículos de motor –

DATOS DE CONTACTO

Empresa:
Maquinarias Santiago Ltda

Actividad económica :
Comercio al Por Mayor y Menor; Rep. Vehículos Automotores/Enseres Domésticos

DESCRIPCIÓN

Mecánico diesel o bencinero, ojala con experiencia en sistemas eléctricos o neumática, desarrollara labores de mantenimiento y reparaciones menores, con licencia de conducir y disponibilidad para viajes a otras regiones

Metropolitana, La Cisterna | 29/08/2017 | $ 400.000 - 500.000 | Jornada Completa

REQUISITOS SOLICITADOS

Nivel educacional:
Educación media completa

Experiencia:
12 meses

Tipo educación media:
- Educación Técnico Profesional

Otros requisitos solicitados:

Licencias de conducir:
- CLASE B - Vehículos motorizados de tres o cuatro ruedas para transporte particular (automóviles, camionetas, furgones, furgonetas, etcêtera).

CARACTERÍSTICAS

Tipo de contrato:
Contrato indefinido

Cargo ofrecido:
Profesional

Ilustración 2.2: Ejemplo de Aviso de trabajo publicado en la BNE.
Fuente: Plataforma laboral BNE.

En estos avisos se incluye información estandarizada y atractiva para su inclusión en el estudio realizado en esta memoria, en la tabla 2.2 se describen estos datos.

ID	DATO	DESCRIPCIÓN
1	Título	Área de texto libre con el título de la oferta de trabajo.
2	Vacantes	Cantidad entera de vacantes que tiene la oferta
3	Ocupación	Ocupaciones que abarca la oferta, normalizada por una lista de ocupaciones de la BNE.
4	Empresa	El nombre de la empresa que publica la oferta.
5	Actividad Económica	Tipo de actividad económica a la que pertenece la empresa.
6	Descripción	Área de texto libre donde se describe el trabajo a realizar, sus características y las competencias necesarias para el cargo.
7	Región y Comuna	Locación estandarizada donde se ejecuta el trabajo.
8	Fecha de publicación	La fecha estandarizada de publicación del aviso.
9	Remuneración	El rango de remuneración que contempla la oferta laboral.
10	Tipo de Jornada	Tipo de jornada laboral estandarizada.
11	Nivel Educacional	Nivel educacional estandarizado requerido por la empresa.
12	Experiencia	Tiempo de experiencia que se exige a los postulantes.
13	Tipo de Educación Media	Tipo de educación exigida a los postulantes.
14	Otros Requisitos	Otros requisitos necesarios para la postulación como licencias de conducir, cursos, certificados, etc.
15	Tipo de Contrato	El tipo de contrato estandarizado ofrecido.
16	Tipo Cargo Ofrecido	Tipo de cargo estandarizado.

Tabla 2.2: Datos dentro de un aviso de trabajo publicado en la BNE.
Fuente: Elaboración propia.

De los datos asociados a un aviso de trabajo se destaca en primer lugar el campo nº 6, que corresponde a la descripción del aviso laboral, esta es un área de texto libre donde se permite la libertad de escribir lo que desee cada persona o empresa que ofrezca un trabajo. En ella generalmente se declaran las características que tendrá el trabajo, algunas de las funciones que deberá realizar en la actividad y, aún más crucial, se describen las características deseadas del postulante junto a las competencias necesarias para su ejecución. Además, el campo nº14, también es incluido en el análisis, ya que es un área de texto libre donde los usuarios especifican otros requisitos que expliquen las necesidades del puesto laboral ofertado. Sobre estos campos se aplican las herramientas de text mining para la detección de competencias laborales.

Luego hay una serie de campos o datos en el aviso de trabajo que son normalizados por parte de la BNE, desde su formulario para la publicación de dicho aviso el sistema le ofrece ciertas opciones predeterminadas al usuario para completar la inscripción. Estos campos

corresponden a los siguientes: Ocupación, Actividad Económica, Vacantes, Nivel Educacional, Tipo de Educación Media, Región y Comuna, Fecha de publicación, Tipo de Jornada, Tipo de Contrato y Tipo de Cargo Ofrecido. Estos campos permiten la segmentación y agrupación de los resultados, tomando esos datos como dimensiones para el estudio a través de la plataforma dispuesta.

2.2. Minería de Texto[5].

Para definir lo que es minería de texto, desde ahora text mining, es necesario empezar mencionando lo que es la minería de datos o data mining. Data mining corresponde al proceso de aplicar algoritmos para encontrar patrones no triviales dentro de información previamente preparada y seleccionada. Estos patrones son interpretados, luego entregan nueva información que potencialmente tendrá algún tipo de utilidad. [4]

Dentro del campo del Data Mining, existe un área que estudia el análisis especifico de información en formato de texto no estructurado, esta área es denominada como Text Mining. La información puede provenir desde documentos como artículos de diarios, publicaciones científicas, libros y, con mayor protagonismo hoy en día, desde páginas web.

2.2.1. Pre procesamiento de datos.

El pre procesamiento de datos corresponde a un conjunto de técnicas que se aplica a la información para reducir su variabilidad, normalizarla y limpiar los datos de anomalías en el lenguaje utilizado u otros factores. Este proceso es necesario para preparar los datos para una posterior aplicación de métodos y algoritmos de text mining.

A continuación, se presentan las técnicas más utilizadas en este proceso.

Tokenización.

Tokenización corresponde al proceso de particionar un documento de texto, separando su secuencia caracteres mediante la identificación de los límites de cada palabra, donde ésta empieza y donde termina. Cada componente resultante es denominado **Token**. A veces se pueden incluir métodos que identifiquen distintas formas de referirse a una palabra particular y la normalicen, por ejemplo, frente a la presencia de palabras como *"Usted", "Ud.", "ud."*. [5] La agrupación de estos tokens en un vector conforman lo que es comúnmente llamado bolsa de palabras.

[5] *Minería de datos también es referido como Text Mining.*

A veces es recomendable que en esta etapa se aplique una normalización al texto, eliminando tildes, puntuaciones o disminuyendo todas las letras a minúsculas.

Eliminación de Stopwords.

Luego de segmentar una serie de documentos o tokenizarlo, es probable terminar con un set de miles de palabras, donde se pueden encontrar palabras que no son muy relevantes semánticamente como lo son algunos pronombres, artículos o preposiciones, tales como, *"a", "por", "el", "de",* entre otros.

Entonces se define una lista de palabras sin real aporte a los conceptos que se busca estudiar o sin real aporte semántico. Estas palabras se denominan **stopwords,** y son eliminados del set de palabras en los datos.

Es natural deducir que esta lista será diferente dado el lenguaje utilizado y también el contexto, un análisis de oraciones puede no eliminarlos, mientras que otro en busca de conceptos en artículos científicos sí. [4]

Stemming.

El lenguaje permite muchas variaciones morfológicas de una palabra, podemos ejemplificar con la palabra, "*Calcular*", de ella surgen variantes como *"calculado", "cálculo", "calculadora", "calculable"* y otras.

La aparición de estas variaciones en los datos y su trato como individuales puede dar indicaciones erróneas de la significancia del concepto dentro del documento. Entonces aparece el proceso de stemming o derivación en español, donde todas estas variaciones se reducen a su raíz, en el ejemplo anterior la raíz sería *"calcul"*. Este método reduce significantemente la cantidad de palabras utilizadas por los algoritmos y mejora su desempeño, pero hay que tener precaución de no eliminar conceptos valiosos dentro del contexto de investigación. [4]

En particular se implementó la metodología para el español desarrollada por Dr. Martín Porter [27] [28]. A continuación, se explica este método:

- Las vocales son: *{ a ; e ; i ; o ; u ; á ; é ; í ; ó ; ú ; ü }.*

- Se definen regiones R1 y R2 en cada palabra.

    - R1: Es la región luego de la primera letra no vocal que sigue a una vocal, o si no, es la región nula después de la palabra en caso de que no exista una palabra no vocal.

    - R2: Es la región siguiente a la primera no vocal que siga a una vocal dentro de R1, o es la región nula al final a la palabra si no existe tal letra no vocal.

- Se define la región RV:
    - Si la segunda letra es una consonante, es la región que sigue después de próxima vocal.
    - Si las dos primeras letras son vocales, es la región luego de la siguiente consonante.
    - Si las dos primeras letras son consonantes y vocal, en ese orden, es la región siguiente a la tercera letra.
    - Si no hay ni uno de esos casos RV está al final de la palabra.

Pasos:

Siempre se realiza el paso 0 y 1.

0. Pronombre adjunto:
    a. Se busca la coincidencia más larga entre los siguientes *{ me ; se ; sela ; selo ; selas ; selos ; la ; le ; lo ; las ; le s; los ; nos }*
    b. Se elimina si viene luego de los términos en RV:
        i. *{ iéndo ; ándo ; ár ; ér ; ír }*. La eliminación también implica la remoción de la tilde.
        ii. *{ ando ; iendo ; ar ; er ; ir}*.
        iii. *{yendo}* luego de una *"u"*. Yendo debe estar dentro de RV, pero la *"u"* puede estar fuera.
1. Remoción estándar del sufijo:
    a. Se busca el término presente más largo entre los siguientes *{ anza ; anzas ; ico ; ica ; icos ; icas ; ismo ; ismos ; able ; ables ; ible ; ibles ; ista ; istas ; oso ; osa ; osos ; osas ; amiento ; amientos ; imiento ; imientos }* y se elimina si se encuentra en R2.

b. *{ adora ; ador ; ación ; adoras ; adores ; aciones ; ante ; antes ; ancia ; ancias }* Se eliminan si están en R2. Si están antecedidas por *"ic"* se elimina si esta en R2.

c. *{ logía ; logías }* Se reemplazan con *"log"* si esta en R2.

d. *{ ución ; uciones }* Se reemplazan con *"u"* si esta en R2.

e. *{ encías ; encías }* Se reemplazan con *"ente"* si esta en R2.

f. *"amente"* se elimina si esta en R1. Si es antecedida por *"iv"* y seguida por *"at"* se elimina si esta en R2, u otro caso, si esta antecedida por *{ os ; ic ; ad }* se elimina si esta en R2.

g. *"mente"* se elimina si esta en R2, si esta antecedido por *{ ante ; able ; ible }* se elimina en R2.

h. *{ idad ; idades }* Se eliminan si están en R2. Si esta antecedido por *{ abil ; ic ; iv }* Se elimina si esta en R2.

i. *{ iva ; ivo ; ivas ; ivos }* Se eliminan en R2. Si esta antecedido por *"at"* Se elimina en R2.

2.

a. Se realiza si el final de la palabra no fue removido por el paso 1.

    i. Sufijos de verbos que comienzan con *"y"*. Se busca el termino presente en RV más largo entre los siguientes *{ ya ; ye ; yan ; ye ; yeron ; yendo ; yo ; yó ; yas ; yes ; yais ; yamos }* Se eliminan si los antecede *"u"*.

b. Se realiza si el paso 2a se hizo, pero falló en eliminar el sufijo.

    i. Sufijos de otros verbos. Se busca el sufijo más largo presente en RV entre los siguientes *{ en ; es ; éis ; emos }* y se elimina si lo antecede *"gu"* o *"u"*.

    ii. Se eliminan los siguientes términos *{ arían ; arías ; arán ; arás ; aríais ; aría ; aréis ; aríamos ; aremos ; ará ; aré ; erían ; erías ; erán ; erás ; eríais ; ería ; eréis ; eríamos ; eremos ; erá ; eré ; irían ; irías ; irán ; irás ; iríais ; iría ; iréis ; iríamos ; iremos ; irá ; iré ; aba ; ada ; ida ; ía ; ara ; iera ; ad ; ed ; id ; ase ; iese ; aste ; iste ; an ; aban ; ían ; aran ; ieran ; asen ; iesen ; aron ; ieron ;*

ado ; ido ; ando ; iendo ; ió ; ar ; er ; ir ; as ; abas ; adas ; idas ; ías ; aras ; ieras ; ases ; ieses ; ís ; áis ; abais ; íais ; arais ; ierais ; aseis ; ieseis ; asteis ; isteis ; ados ; idos ; amos ; ábamos ; íamos ; imos ; áramos ; iéramos ; iésemos ; ásemos }.

3. Siempre se realiza el paso 3. Sufijo residual.

    a. Se busca el sufijo presente más largo entre *{ os ; a ; o ; á ; í ; ó }* y se elimina si esta en RV.

    b. *{ e ; é }* Se eliminan si están en RV. Si están antecedidos por *"gu"* con la *"u"* dentro de RV, se elimina la *"u"*.

4. Final. Se eliminan las tildes.

Lematización.

Lematización es un proceso similar al stemming, pero con una diferencia fundamental, mientras que stemming es una forma heurística de reducir la cantidad de palabras llevándolas a su raíz, la lematización en cambio busca agrupar palabras que refieran al mismo concepto semántico, por ejemplo, *"miré"* y *"vi"* mediante stemming se reducen a "*mir*" y "*vi"*, pero el proceso de lematización las reduciría a una forma común como *"mir"* o *"ver"*, esta base común es conocida como el **Lema** de la palabra. [6]

N-gramas.

Ya se explicó anteriormente una forma de representar un documento de texto, como lo es su forma de bolsa de palabras, esta representación ignora el orden de ellas en el documento y teniendo en cuenta que el contexto dentro del lenguaje es vital para su interpretación, se podría estar perdiendo información relevante.

Para solventar este inconveniente es que se implementa los llamados n-gramas para la representación del documento, estos aparte de tokenizar el documento también consideran el vecindario de cada componente. Si bien su aplicación puede enriquecer el análisis, conlleva un costo del crecimiento exponencial del set, por lo cual es extraño ver su aplicación más allá de duplas o tripletas entre palabras. Por lo general se recomienda empezar con un análisis de palabras individuales y luego probar si la inclusión de n-gramas mejora los resultados significativamente. [7]

Un ejemplo de n-gramas es el siguiente, se considera la frase "Buenos días, capitán." Si se le aplica un n-grama con $n = 2$, dicho de otra forma, un bi-grama, el conjunto de palabras, c, sería la siguiente [41]:

$$c = \{\, Buenos\,;días\,;capitán\,;\ Buenos.días\,;\ días.capitan\,\}$$

2.2.2. Taxonomía de competencias laborales.

En esta memoria se tiene como uno de sus objetivos principales identificar competencias laborales dentro de un aviso de trabajo. Para ello es necesario tener previamente definido un set o catálogo de competencias laborales que representen las habilidades, conocimientos y aptitudes necesarias para ejercer funciones laborales dentro del mercado o la industria general, en este caso la chilena.

Para hacer uso de este catálogo o conjunto de competencias es necesario definir una taxonomía de estos conceptos. Una taxonomía es un Sistema o Estructura Organizado de Información. [8]

Respecto a taxonomías de competencias laborales se encuentran esfuerzos por distintas organizaciones que investigan y confeccionan sus propias taxonomías laborales para distintos propósitos, por ejemplo, la ESCO: European Skills/Competences, Qualifications and Ocupations, actualiza constantemente una taxonomía en forma de tesauro[6], esto quiere decir que mantiene relaciones y jerarquía entre sus elementos. Esto con fines de apoyar en lograr los objetivos de Estrategia Europa 2020, que abarca un crecimiento sostenido en los empleos para la comunidad europea. [9]

En un contexto nacional existe la institución ChileValora: Comisión Sistema Nacional de Certificación de Competencias Laborales, el cual elabora un catálogo de con cerca de 3.000 competencias laborales.

A partir de este catálogo ChileValora se encarga de aprobar a centros certificadores, llevar un registro de los trabajadores certificados, elaborar perfiles, además de servir como apoyo a la elaboración de cursos de capacitación, entre otras labores. [10]

Si bien la publicación de estas competencias laborales es declarada como un catálogo de competencias laborales, conceptualmente corresponde a una taxonomía en forma de vocabulario controlado, ya que su contenido es decidido bajo cierto protocolo y bajo responsabilidad de un actor, ChileValora.

2.3. Data warehouse.

Un data warehouse se define como una colección de datos orientada a un sujeto, integrada, no volátil y variante en el tiempo para el apoyo de decisiones de los directivos de un área o de decisiones no operacionales. [15]

---

[6] *Dicha taxonomía contiene más de 10.000 competencias, pero no se encontraba disponible al público al momento de la realización de esta memoria.*

Se decide implementar un data warehouse como sistema para almacenar la información generada a través de la detección de competencias laborales en avisos de trabajos de la BNE, de esta forma se permite el estudio de los datos a nivel agregado dependiendo de los aspectos a los que se le quieran dar enfoque.

Para un mejor entendimiento de la arquitectura básica de un data warehouse, esta queda plasmada en la ilustración 2.3

2.3.1. Fuente del sistema operacional.

En primer lugar, se encuentra la fuente del sistema operacional, que corresponde al sistema que genera los datos utilizados en el sistema general.

Esta fuente debe contar con un desempeño de procesamiento y disponibilidad adecuados a los requerimientos de extracción de información que se planeen realizar.

En este trabajo de título corresponderá a la fuente de información que contiene los avisos de trabajo junto a las competencias presentes en él.

2.3.2. Data Staging Area.

La segunda componente de la arquitectura es el Data Staging Area, traducido como etapa de preparación de los datos, en esta componente se almacenan y realizan pequeños procesos a los datos a forma de preparación para su almacenamiento en el modelo final del data warehouse. Los procesos de transformación posteriores de los datos que incluyen su transformación, extracción y carga se denomina proceso ETL, por sus siglas en inglés, “Extraction”, “Transformation” y “Load”. [42]

En esta componente los datos son extraídos desde la fuente de datos, luego transformados y procesados, puede incluir una limpieza de los datos, combinación de distintas fuentes de datos, eliminación de datos duplicados, entre otros.

2.3.3. Modelo Estrella

Dentro del sector de almacenamiento de datos del data warehouse, se usa un modelo de datos relacional, en el cual la entidad principal se denomina tabla Fact, en ella van los resultados que se mostraran luego a los clientes y usuarios del sistema, y tiene relacionado las llamadas dimensiones. Las dimensiones son características propias de los resultados almacenados, estas sirven para segmentar o agrupar la información que se mostrará en los reportes que entregará el data warehouse. [16]

2.3.4. Arquitectura del Data Warehouse.

A continuación, se detalla la arquitectura tecnológica dentro de un data warehouse. Una representación de esta se encuentra en la Ilustración 2.3. Al principio el usuario ingresa a través de un navegador web a la plataforma del data warehouse, en este posee herramientas para personalizar y generar distintos reportes o tableros con la información almacenada en la plataforma.

Para generar esos reportes y tableros la plataforma traduce los requerimientos del usuario a consultas MDX. El lenguaje MDX fue desarrollado para consultas en un ambiente de procesamiento analítico en línea, desarrollado por Microsoft. [46]

Estas son recibidas por el motor Mondrian, dicho motor se encarga de procesar el análisis pedido por el usuario y recibido en formato MDX.

Luego el motor, utilizando el esquema Mondrian como guía, genera una consulta SQL que responde a la petición del usuario. El esquema Mondrian es una forma de definir un modelo lógico de datos.

Cabe destacar que Mondrian es una tecnología de software libre en el cual se implementa el procesamiento de un data warehouse. Como esta existen más opciones disponibles.

Finalmente, se ejecuta la consulta SQL al modelo estrella y la información de vuelta es tomada por el motor Mondrian para generar los reportes y ser entregados al usuario en su interfaz desde el navegador web.

Las componentes de esta arquitectura son presentadas con mayor detalle en el siguiente capítulo.

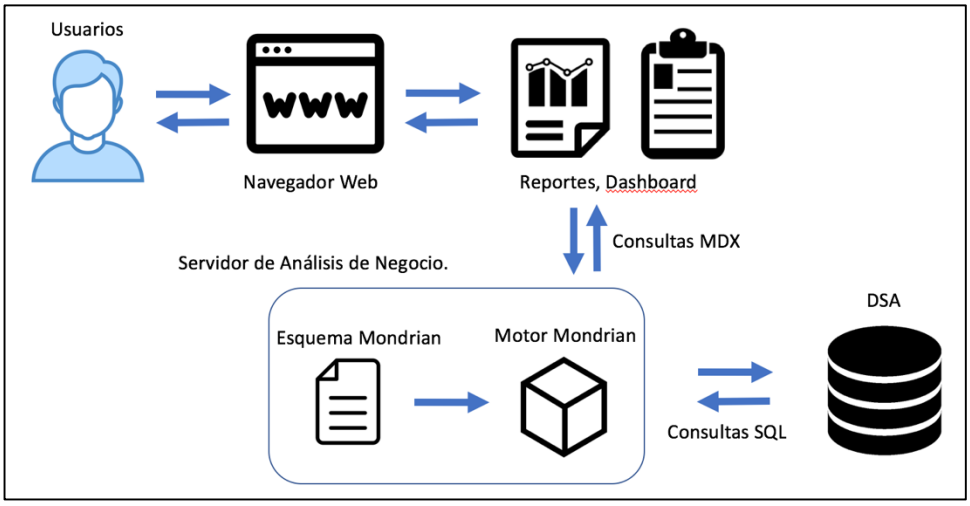


Ilustración 2.3: Esquema Básico de la Arquitectura de un Data Warehouse.
Fuente: Elaboración propia.

Capítulo 3

Arquitectura de la Plataforma de Análisis e Identificación de Competencias Laborales

La estructura general del sistema busca responder a los requerimientos de integrar las fuentes de información de la BNE con las competencias laborales definidas por ChileValora usando la herramienta de detección de competencias laborales

Con esto busca poder llevar los resultados del procesamiento a un usuario final, que por medio de las herramientas provistas por el sistema data warehouse podrá realizar un estudio en profundidad de las competencias laborales encontradas en los avisos de trabajo de la BNE.

En este capítulo se dará una explicación general a los componentes de la arquitectura, y en capítulos posteriores se detallará las características de cada uno. Una vista general de la arquitectura puede ser encontrada en la Ilustración 3.1.

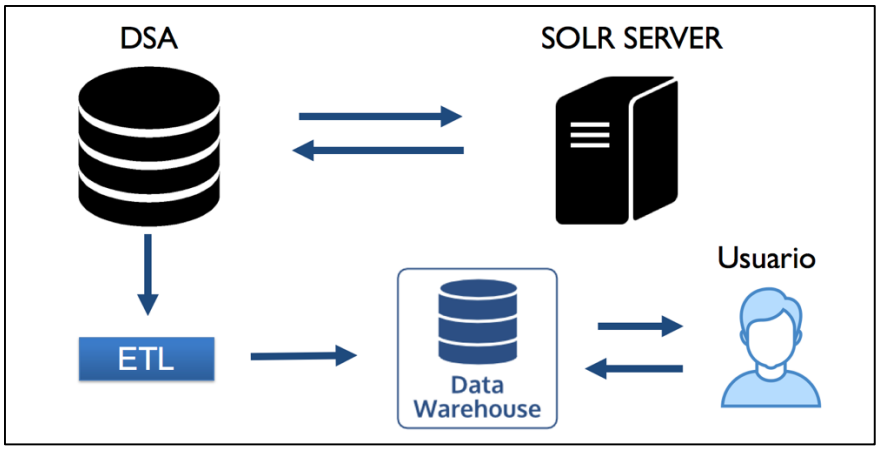


Ilustración 3.1: Arquitectura general del sistema.
Fuente: Elaboración Propia.

Las componentes que conforman la arquitectura son las siguientes.

1. SOLR Server.
2. DSA: Data Staging Area.
3. Proceso ETL: Extraction, Transformation and Load.
4. Data Warehouse.

3.1. SOLR Server.

Conocido como Apache SOLR, es un motor de búsqueda de código abierto, desarrollado por Apache Software Foundation (ASF), es una de las organizaciones sin fines de lucro más reconocidas en el ámbito de proveer soluciones y softwares para el bien público de forma gratuita. [29] [30]

Entre sus características se encuentra la posibilidad de implementar un buscador para sus colecciones que escala manteniendo su desempeño a pesar de un aumento en el tamaño de la colección.

Este servidor provee herramientas e infraestructura de software que permite realizar búsquedas complejas dentro de grandes colecciones de información. Incluye herramientas para aplicar técnicas de Text Mining, como en el pre procesamiento de texto, tanto para la información almacenada como para las consultas u objetos que se buscan en el servidor.

Así mismo, integra herramientas para hacer búsquedas dentro de la colección almacenada permitiendo una gran configuración que defina reglas para las búsquedas.

Dicha búsqueda asigna un puntaje a los documentos que coincidan en la búsqueda, luego prioriza los resultados según su puntaje y entrega la información en un formato especificado. Este formato puede ser amigable para el usuario o utilizado para la comunicación entre máquinas.

Esta arquitectura cumple la función de almacenar y pre procesar los avisos de trabajos de la BNE y pre procesar las competencias laborales definidas por ChileValora. Para luego realizar una búsqueda inteligente que permita encontrar dichas competencias dentro de los avisos laborales.

3.2. DSA: Data Staging Area.

Descrito anteriormente en el Marco Teórico su funcionalidad dentro de la arquitectura no se aleja mucho de esa caracterización. En esta componente, se almacena la información correspondiente a las competencias laborales, a los avisos de trabajo y sus características, pero con mayor importancia se almacenan las coincidencias encontradas entre competencias laborales y avisos de trabajo, junto a su puntaje obtenido. Toda esta información será luego

procesada y cruzada con las distintas fuentes de información para ser incluida dentro del data warehouse.

Su implementación fue llevada a cabo dentro de un entorno provisto por XAMPP, una distribución gratuita de Apache Friends. Esta permite fácilmente la implementación de un servidor web apache.

Por otro lado, utiliza MariaDB para la administración de bases de datos relacionales, su elección se debe a la inclusión de un módulo para administrar base de datos y facilidad de uso para el propósito destinado.

MariaDB es una base de datos de software libre con más de 12 millones de usuarios en el mundo, esta corporación fue creada por los mismos fundadores de la base de datos MySQL [43].

3.3. Proceso ETL.

El proceso ETL, extracción, transformación y carga de datos, tiene como función preparar la información almacenada en el DSA para su carga al sistema data warehouse.

En el caso particular de esta memoria debe hacer un cruce de información entre los avisos de trabajos, la información de estos y de las empresas que publicaron dichos avisos para conformar las dimensiones del modelo estrella del data warehouse

Luego debe hacer realizar una transferencia de esta información al modelo final, relacionándola con las coincidencias encontradas y el puntaje asignado a dicha coincidencia.

Finalmente, la información quedará almacenada bajo la estructura relacional del modelo estrella en el mismo motor de búsqueda que fue descrito para el DSA anteriormente. Puediendo ser accedido a través de las consultas generadas en el data warehouse.

3.4. Plataforma de BI o Data warehouse.

La plataforma de análisis escogida para el sistema data warehouse que analiza y estudia las competencias laborales demandadas por los avisos de trabajos es “Saiku Community Edition” un software de código abierto, fundado en 2008 bajo el nombre “Pentaho Analysis Tool” y renombrado en 2010 como Saiku.

Saiku provee una plataforma configurable, de fácil uso, para el análisis de datos e incorporación de sus distintas dimensiones al estudio. Es una tecnología OLAP, “On-line analytical Processing”, o en español, procesamiento analítico en línea. Centrado en análisis multidimensional de manera interactiva.

Particularmente Saiku está desarrollado para conectarse con distintos servidores de análisis OLAP, por defecto, utiliza Mondrian, un servidor o motor OLAP de código abierto basado en

el lenguaje de programación Java. Este provee las herramientas y metodologías para hacer los análisis y consultas, mientras que Saiku es la interfaz que conecta al usuario con el sistema.

Permite al desarrollador configurar los elementos que se medirán en el data warehouse, las dimensiones o características que rodean a cada elemento, su acceso es vía un navegador web y posee características amigables para el usuario, como lo es el "drag and drop" arrastrar y soltar, para seleccionar los atributos por los cuales medir las competencias laborales. [36]

Finalmente, Saiku ofrece una REST API, un tipo de interfaz de programación de aplicaciones que habilitan la comunicación entre servicios o programas externos con el servidor Saiku del data warehouse para hacer peticiones o generar reportes sin la interacción de un usuario.

3.5. Esquema Mondrian.

La arquitectura del data warehouse necesita de un mapa o guía de que preguntas puede responder, en cuanto a que medidas o métricas calcula el sistema, por cuales dimensiones y jerarquías puede analizarse dichas medidas, desde que entidades o tablas particulares provienen esos elementos en el modelo estrella y como se relacionan entre sí. [37] [38]

Dicho mapa o guía se denomina esquema, este define una base de datos multidimensional, formando un modelo lógico, este se compone de:

- Cubos: Sub segmentos específicos de la totalidad de la información que analizan cierto aspecto específico del negocio o sistema.
- Jerarquías: Un orden anidado de acuerdo a las dimensiones o características que agrupan la información, como lo puede ser país, provincia, ciudad donde fueron publicados los avisos laborales.
- Modelo físico: Indica en qué tablas o entidades del sistema de administración de la base de datos, son la fuente de los elementos anteriormente descrito y como estos se relacionan entre sí.

Este esquema se utiliza para que el sistema data warehouse genere consultas en lenguaje MDX, expresiones multidimensionales, que es una forma de consultar métricas especificas al sistema diseccionando por las dimensiones especificadas.

Finalmente, el esquema se traduce a un archivo XML donde se especifican todos los aspectos descritos anteriormente para luego ser cargado al servidor del data warehouse que le permite hacer consultas a la fuente de los datos y al usuario hacer distintos análisis.

Capítulo 4

Diseño e implementación del sistema de detección de competencias laborales

En el presente capitulo se detalla la implementación de las componentes que conforman el sistema de detección de competencias laborales, desde la extracción de los avisos de trabajo de la BNE, recorriendo su integración con el repositorio de competencias laborales, la configuración y levantamiento del servidor de búsqueda, la implementación de la herramienta que genera las coincidencias, finalizando en el almacenamiento de los resultados en el DSA.

4.1. Repositorio de Avisos laborales y empresas.

La fuente original del repositorio de avisos laborales para esta memoria consta de una base de datos bajo el dominio de la Universidad de Chile, esta cuenta con 690.525 registros de avisos de trabajos cargados en el sistema de la BNE, entre los años 2011 y 2016. Estos datos fueron extraídos directamente del sistema operacional de la BNE, por lo que sus datos no han sido limpiados aun, existiendo registros erróneos dentro del repositorio.

La estructura de estos avisos laborales se encuentra detallados en el Marco Teórico, junto con un ejemplo y una explicación de los campos que lo conforman. En la Ilustración 2.2 y la Tabla 2.2 respectivamente.

Es necesario poder almacenar esta información en un entorno local para su procesamiento directo. Debido a que, si no se poseen los avisos en un ámbito local, la comunicación entre las distintas máquinas se convertiría en un cuello de botella para el procesamiento siendo afectado por factores como la conexión, su estabilidad y la velocidad de ésta.

Para satisfacer la necesidad recientemente descrita, se realizó el siguiente proceso: en primer lugar, se utilizó el programa Microsoft Excel junto a una conexión mediante ODBC, Open Database Connection, una interfaz estándar para comunicarse con un sistema de administración de base de datos sin importar el sistema donde se encuentre implementado este y su sistema operativo. Mediante esta herramienta se realizaron consultas para extraer en primer lugar las entidades del sistema descritas en la Tabla 4.1. Todas las extracciones fueron almacenadas en archivos de formato Microsoft Excel.

ENTIDAD	DESCRIPCIÓN	EJEMPLOS
ÁREA	Área laboral estandariza a las cuales pertenece el trabajo ofrecido.	Salud, Gastronomía, Publicidad.
CARGOS GENERALES	Tipo de cargo estandarizado del cargo.	Gerente, Ejecutivo, Supervisor.
DISPONIBILIDAD	Tipo de disponibilidad horaria estandarizada solicitado por la empresa.	Fulltime, Part-time, Por turnos.
REGIÓN	La región desde donde se ofrece la oferta laboral.	Metropolitana, V Región de Valparaíso.
NIVEL COMPUTACIONAL	Nivel computacional estandarizado necesario para el trabajo.	Usuario, Básico, Técnico, Experto.
CARRERAS	Carrera universitaria, técnica u oficio asociado al puesto de trabajo.	Ingenierías varias, Panadero, Paramédico.
GRADO ESCOLAR	Grado de escolaridad estandarizado requerido por la empresa.	Básica, Media, Universitaria.
NUMERO DE EMPLEADOS	Rango correspondiente a la cantidad de empleados de la empresa.	1-10, 501-1.000, Más de 5.000
SITUACIÓN DE ESTUDIOS	Situación de estudios solicitado a los postulantes.	En curso, Graduado, Egresado.
PERFIL DE CANDIDATO	Perfil del postulante en relación al nivel educacional.	Estudiante, Estudios Superiores, Oficios.
ACTIVIDAD EMPRESARIAL	Tipo de actividad empresarial de la empresa ofertante.	Construcción, Imprenta, Hotelería.

Tabla 4.1: Entidades extraídas.
Fuente: Elaboración propia.

Luego se extrajo, mediante el mismo método, la entidad correspondiente a las empresas y sus características ingresadas en el sistema. Para publicar un aviso laboral, el usuario individual o la empresa debe registrarse en la plataforma de la BNE. En total hay registros de 70.367 empresas distintas en el sistema.

Para los avisos laborales se realizó una modificación al método de extracción. Debido a la cantidad de registros que conforman el repositorio de avisos laborales se debió limitar la extracción a 50.000 registros por tramo, hasta completar los cerca de 700.000 avisos laborales. Estos aparte de incluir los campos descritos en la Tabla 2.2 poseen campos que no son de acceso público y que también son de uso del sistema.

Finalmente se extrajo una serie de entidades que son tablas relacionales entre las empresas o avisos y las entidades de la Tabla 4.1, que reflejan relaciones “n es a m” o “n-m” entre éstas. Esto quiere decir que un registro de una empresa, por ejemplo, puede tener asociado distintas actividades empresariales y de forma inversa, una actividad puede tener asociada muchas empresas. Estas entidades sirven de intermediario entre dimensiones y registros de la empresa.

4.2. Repositorio de Competencias Laborales.

Las competencias laborales dentro del sistema completo desarrollado en esta memoria son un insumo clave dentro de éste, pero al mismo tiempo son un insumo reemplazable y tolerante a cambios, por lo que la elección de la colección de competencias puede cambiar, pero el funcionamiento y su implementación dentro del sistema no variará mucho.

De todas formas, la mejor colección de propuestas laborales para el desarrollo de esta memoria corresponde a la elaborada por ChileValora, esta se creó sobre el mercado nacional, sobre la base a los trabajos desarrollados en este, reflejando las necesidades de los puestos de trabajo en cuanto a competencias se refiere. Al mismo tiempo la BNE contempla avisos laborales del mercado chileno. Teniendo una relación directa entre competencias laborales y avisos de trabajo debido a que surgen dentro del mismo mercado y contexto.

ChileValora posee como un recurso público un catálogo de competencias laborales que consta de 2.839 competencias laborales, cada una de ellas es conformada por un código de Unidad de Competencia Laboral (UCL) que identifica a la competencia, el nombre de este que funciona también como un descriptor de la competencia, a su vez incluye el código y el nombre del perfil laboral al cual pertenece la UCL, y finalmente el sector industrial al que pertenece el perfil laboral.

Esta colección está disponible para su descarga en formato Microsoft Excel en el portal de ChileValora [31].

La integración del catálogo de competencias laborales al sistema local, se implementó mediante un programa propio, desarrollado en lenguaje Java, el cual se encargaba de leer una versión del catálogo en formato separado por comas (.csv) línea por línea, identificando los campos de interés, y mediante una conexión al DSA, realizando la inserción de datos. Quedando como resultado las cerca de 3.000 competencias dentro de la base de datos en el DSA, disponibles para su lectura mediante simples consultas en formato SQL.

4.3. Motor de Búsqueda: SOLR.

El motor de búsqueda SOLR debe pasar por distintas etapas para poder utilizar sus funcionalidades y cumplir su rol dentro de la arquitectura. Este servidor tiene la capacidad de almacenar distintas colecciones de datos, con diferentes configuraciones.

Estas colecciones pueden tener una configuración predefinida para el esquema de datos a utilizar definida por el desarrollador, en esta se especifican los tipos de datos a almacenar, los campos que existirán, y con mayor importancia, los procesos y transformaciones que sufrirán los datos insertados para su posterior indexación. La indexación permite una búsqueda rápida dentro de los campos de texto u otros que se encuentren en el servidor.

Todas las configuraciones son archivos en formato Extensive Markup Language (.xml) que pueden ser modificados para la generación de colecciones con las especificaciones del desarrollador.

4.3.1. Configuración de la colección.

La colección desarrollada para esta memoria se nombró “Avisos” y consta de dos elementos para cada registro:

- AVISOID: Corresponde al identificador del aviso laboral, fue obtenido directamente del repositorio de avisos laborales de la BNE y no fue modificado previo a su inserción. Este campo es indexado por el sistema.

- AVISOCUERPO: Corresponde a una suma de los campos de texto Descripción (ID 6) y Otros Requisitos (ID 14) que se encuentran en los avisos de trabajo de la BNE, y están descritos en la Tabla 2.2 dentro del capítulo Marco Teórico.

Otro aspecto que se debe configurar luego de definir los campos de la colección, son el tipo de dato que se usara en cada campo junto a las características de éste.

Para el AVISOID se utilizó un tipo de dato llamado “int” haciendo referencia a su naturaleza de ser un número entero o “Integer" en inglés. Fue configurado para que el motor de búsqueda lo indexe y almacene su valor luego de ingresar un nuevo registro, pero cabe destacar que es fijado como una llave única, es decir, ni un registro puede tener el mismo id, así se evitan posibles duplicados. Este campo no posee mayor complejidad ni necesidad de análisis.

A continuación, se describe la configuración del campo AVISOCUERPO, que es vital para el funcionamiento del sistema desarrollado y para la herramienta de detección de competencias laborales.

El campo AVISOCUERPO fue asociado a un tipo de dato nombrado “Texto general”, en primer lugar, se dio instrucción para que el valor de este campo fuera almacenado e indexado para su búsqueda posterior. También se configuró el pre procesamiento que debía pasar cada registro antes de ser ingresado al servidor, todos estos procesos utilizan herramientas provistas por defecto en el servidor, pero deben ser incluidas manualmente. El proceso se describe a continuación:

1. Minúscula: Todas las letras en mayúsculas del campo son transformados a minúscula.

2. Reemplazo de símbolos: Se transforman todos los caracteres del campo que no pertenecen al conjunto de Latín Básico de la codificación de caracteres ASCII a su representación en dicha codificación.

3. Eliminación de stopwords: Se eliminan las stopwords del campo de texto, estas son un insumo provisto por el desarrollador en un archivo llamado “stopwords” y en formato de texto plano (.txt). Las stopwords en español fueron definidas por el sitio web Snowballstem [35] y adjuntadas en el Anexo A.

4. Stemming: Los términos dentro del campo de texto son transformados a su raíz principal, esto mediante un algoritmo llamado Snowball.

5. N-gramas: Este proceso toma los términos resultantes de los términos anteriores y va generando nuevos términos mediante la agrupación de hasta 3 términos consecutivos.

6. Tokenizador: Finalmente cada termino resultante se separa y se identifica, cada uno de estos es un Token. Estos son los que se indexan para permitir una rápida búsqueda dentro de toda la colección.

Para un mejor entendimiento de este proceso se ejemplifica con un aviso ficticio y se muestra paso a paso su transformación.

Ejemplo: *"Se necesita camarera para restaurante MIRESTAURANT, con buena atención al cliente e impecable higiene."*

1. Minúscula: *"se necesita camarera para restaurante mirestaurant, con buena atención al cliente e impecable higiene."*

2. Remplazo de símbolos: *"se necesita camarera para restaurante mirestaurant con buena atencion al cliente e impecable higiene"*

3. Eliminación de stopwords: *"se necesita camarera para restaurante mirestaurant buena atención cliente impecable higiene"*

4. Stemming: *"se necesit camarer par restaur mirestaurant buen atencion client impec higien"*

5. N-gramas: *"se necesit camarer par restaur mirestaurant buen atencion client impec higien se.necesit se.necesit.camarer necesit.camarer camarer.par necesit.camarer.par par.restaur…" Se omiten todas las tuplas que no se explicitaron por comodidad de lectura.*

6. Tokenizador: *[se; necesit; camarer; par; restau; mirestaurant; buen; atención; client: impec: higien: se.necesit; se.necesit.camarer; ...] Nuevamente no se expresan todos los token generados.*

Este resultado final es indexado y almacenado en la colección dentro del servidor para futuras consultas. Todas las configuraciones anteriormente descritas, los campos de la colección, los tipos de datos, los procesamientos que se aplican a los nuevos registros, se ven reflejados en la Ilustración 4.1, que corresponde al archivo de configuración de la colección, dicho archivo permite replicar el esquema de la colección en otras colecciones u otros servidores de búsqueda SOLR.

```
<?xml version="1.0" encoding="UTF-8" ?>
<schema name="custom-schema" version="1.5">
    <fieldType name="int" class="solr.TrieIntField" precisionStep="0" positionIncrementGap="0"/>
    <fieldType name="texto_general" class="solr.TextField" positionIncrementGap="100">
    <analyzer>
            <tokenizer class="solr.StandardTokenizerFactory"/>
             <filter class="solr.LowerCaseFilterFactory"/>
            <filter class="solr.ASCIIFoldingFilterFactory" preserveOriginal="false" />
            <filter class="solr.StopFilterFactory" words="stopwords.txt" ignoreCase="true"/>
            <filter class="solr.SnowballPorterFilterFactory" language="Spanish"/>
            <filter class="solr.ShingleFilterFactory" maxShingleSize="3" outputUnigrams="true"/>
    </analyzer>
    </fieldType>
    <field name="_version_" type="long" indexed="true" stored="true"/>
    <field name="AVISOID" type="int" indexed="true" stored="true" />
    <field name="AVISOCUERPO" type="texto_general" indexed="true" stored="true" />
    <fieldType name="string" class="solr.StrField" sortMissingLast="true" />
    <uniqueKey>AVISOID</uniqueKey>
</schema>
```

Ilustración 4.1: Archivo de configuración schema.xml.
Fuente: Elaboración Propia.

4.3.2. Puntaje de coincidencia.

El servidor y motor de búsqueda SOLR tiene como finalidad brindar herramientas y una arquitectura para realizar exploraciones dentro de una colección de datos e información, para realizar esto existen métodos de comunicación entre máquinas y una interfaz para usuarios para realizar las consultas.

Sin importar el método de las consultas el resultado será un set de registros que coincidan con la búsqueda realizada además de un puntaje asociado a cada registro.

Este puntaje se construye sobre la base de la presencia de los términos buscados dentro del registro, además del largo relativo del registro con respecto al promedio en la colección y toma en cuenta la presencia de los términos en todos los documentos de la colección. El puntaje busca representar la calidad de la coincidencia, de esta forma, mientras mayor sea el puntaje mayor es la coincidencia entre la consulta y el registro.

A continuación, se presenta y explica el método para calcular el puntaje anteriormente explicado. La función utilizada por defecto en el motor de búsqueda SOLR es **BM25** [32], una función de ranking en el contexto de Búsqueda y Recuperación de Información. Se puede entender como una variación de la función clásica TF-IDF. La fórmula que calcula dicho puntaje se encuentra en la Ilustración 4.2.

$$Se\ considera\ Q\ un\ conjunto\ de\ q_i\ terminos\ dentro\ de\ la\ consulta,$$
$$d, el\ documento\ evaluado,$$
$$\overline{L}, largo\ promedio\ de\ los\ registro\ de\ la\ colección.$$
$$k\ y\ b\ constantes.$$
$$Puntaje(d,Q) = \sum_{i=0}^{n} IDF(q_i) * \left(\frac{tf(d,q_i) * (k+1)}{tf(d,q_i) + k\left(1 - b + b * \frac{|d|}{\overline{L}}\right)} \right)$$
$$Con:$$
$$tf(d,q_i) : Term\ Frecuency.$$
$$El\ nº\ de\ veces\ que\ q_i\ se\ encuentra\ dentro\ del\ documento\ d.$$
$$IDF(q_i) = \log\left(\frac{N - n(q_i) + 0.5}{n(q_i) + 0.5}\right); Inverse\ Document\ Frecuency$$

Ilustración 4.2: Formula de puntaje BM25.
Fuente: Conferencia TREC-3.

Las diferencias entre esta función de puntaje con la clásica TF-IDF radican en diferentes factores. Primero, la frecuencia de términos ($tf(d,q_i)$) tiene un aporte menos pronunciado, luego, la inversa de la frecuencia en los documentos no puede tomar valores negativos, también se incluye un castigo cuando el largo del documento evaluado es mayor que el promedio de la colección. Finalmente, las constantes b y k son parámetros ajustables para la función.

4.3.3. Consultas.

El servidor de búsqueda incluye varias formas para realizar consultas a las colecciones que aloja, estas son las siguientes: [33]

- Standard Query Parser: Tiene énfasis en poseer una sintaxis robusta e intuitiva, pero no muy tolerante a errores de sintaxis.

- DisMax: Proviene de "Maximun Disjunction" o disyunción máxima en español, se enfoca en la búsqueda de frases. Básicamente genera sub consultas a raíz de la consulta original asignándole a cada documento el máximo puntaje obtenido por cualquiera de las sub consultas.

- eDisMax: Significa Extended DisMax, es una versión mejorada de la consulta DisMax. Implementa nuevas funciones, la más importante para esta memoria es la inclusión de considerar una proximidad definida, es decir, considera una separación entre los términos al momento de buscar coincidencias de frases.

Es importante considerar que, al momento de realizar una consulta, el input ingresado, es decir, los términos a buscar (una frase o una competencia laboral) se les aplicara el mismo pre procesamiento que a los registros que ingresados a la colección.

De esta forma, la consulta y la colección poseen el mismo formato y han sido procesados para buscar coincidencias reduciendo sus variaciones lingüísticas mejorando los resultados.

La consulta definida para esta memoria corresponde a eDisMax debido a la posibilidad de permitir movilidad entre los términos de la consulta. Se configuró movimiento de hasta tres términos para hacer coincidencias y además se fijaron los campos que entrega como resultado todos los registros que tengan un puntaje positivo. Luego, se entrega el identificador del aviso, su cuerpo o contenido y su correspondiente puntaje obtenido.

4.3.4. Importación de avisos laborales

En esta sub sección se explica el proceso implementado para la importación de los avisos laborales al servidor SOLR.

1. Se realiza una extracción de la información de los avisos de trabajo mediante una conexión ODBC y la herramienta Microsoft Excel. Se generan extracciones con 50.000 avisos, terminando en 14 archivos del mismo formato Excel.

2. Se aplican macros de Excel a los 14 archivos para eliminar los campos que no se utilizarán, se mantienen el identificador del aviso, campo de texto libre "Descripción" y de "Otros Requisitos" descritos en la Tabla 2.2. Luego la macro

elimina algunos caracteres que interfieren con el formato de archivo separado por comas. Finalmente, los archivos son guardados en el mismo formato (.csv).

3. Luego se implementó un programa propio desarrollado en el lenguaje Java que realiza una lectura de cada archivo con 50.000 avisos en formato separado por comas (.csv). Cada línea corresponde a un registro de los avisos de trabajo.

4. Por cada línea se selecciona la "Descripción" y "Otros Requisitos", se concatenan y se aplican el siguiente filtro para limpiar los datos de avisos erróneos, en primer lugar, el largo de esta variable concatenada debe superar los 11 caracteres y el identificador superar las 8 cifras. En caso de no cumplir esas condiciones se descarta el registro y no se importa a la colección.

5. Para finalizar, se realiza la inserción de cada registro a medida que cumpla las condiciones a través de una conexión http más una consulta SQL.

El resultado de este proceso concluyó con 684.639 avisos filtrados, ingresados e indexados en el motor de búsqueda listos para búsquedas de términos y frases con un tiempo de espera menores a los 3 segundos por consulta.

4.4. Detección de competencias laborales.

Se procede a explicar las componente e implementación del programa encargado de realizar el cruce entre las competencias laborales, los avisos de trabajos almacenados en el motor de búsqueda y el almacenamiento de los resultados en el Data Staging Area.

De forma general el procedimiento corresponde a escoger una cantidad de competencias a buscar, realizar la consulta eDisMax definida anteriormente por cada una de ellas y para cada resultado almacenar la coincidencia en el DSA, guardando el puntaje obtenido, la competencia y el aviso asociado.

El proceso implementado, desarrollado en Java, fue el siguiente:

1. Se define la cantidad de competencias a buscar, estas pueden ser 10, 50 ó 100 competencias.

2. Se define la competencia de partida para el análisis, se realiza una consulta SQL para recopilar las competencias desde el punto de partida y considerando la cantidad estipulada se crea una variable con cada identificador de las competencias junto a su descripción.

3. Por cada competencia recopilada se realiza una consulta eDixMax al motor de búsqueda, utilizando la librería SolrJ (Una API de SOLR para lenguaje Java) [34] y se almacena el resultado en una variable de lista.

4. Por cada registro en el resultado de coincidencias se selecciona el identificador del aviso, el identificador de la competencia y el puntaje obtenido.

5. Mediante una consulta SQL y una conexión http se ingresa la información de la coincidencia como un nuevo registro a la entidad de coincidencias dentro del DSA.

6. Se repite el proceso hasta completar todas las competencias laborales.

Este proceso se representa en la Ilustración 4.3 para una mejor comprensión de la dinámica implementada.

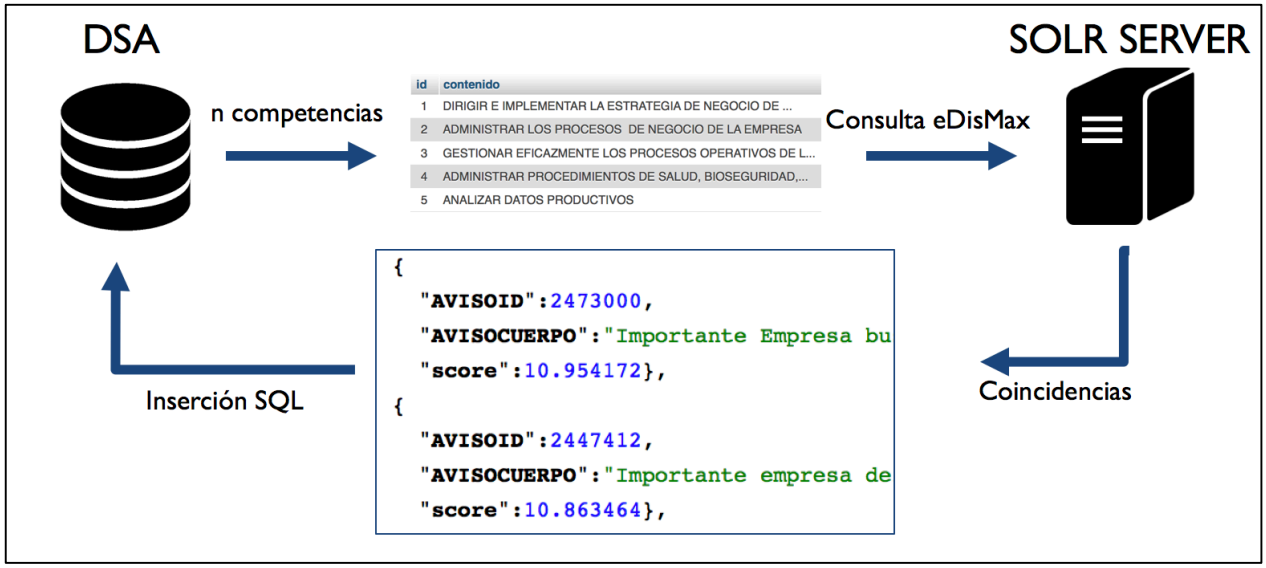


Ilustración 4.3: Diagrama de detección de competencias.
Fuente: Elaboración propia.

4.5. DSA: Data Staging Area.

El soporte tecnológico donde los resultados se almacenan antes de ser insertados en el data warehouse se denomina DSA, este se implementó utilizando MariaDB para la base de datos, y un servidor web apache más un módulo de Phpmyadmin para su administración.

En este soporte se crearon distintas entidades para almacenar los avisos de trabajo, las competencias laborales, las coincidencias generadas por el detector de competencias e información respecto a cada elemento.

Como una competencia puede tener relacionado más de un aviso y en sentido contrario un aviso laboral puede poseer más de una competencia por lo que la relación entre estos corresponde a una relación “n es a m” o “n-m”. El modelo entidad relación implementado se puede observar en la Ilustración 4.4.

En esta base de datos se encuentran almacenados aproximadamente 685.000 aviso labores, 2.150 competencias laborales y más de 400.000.000 de coincidencias entre competencias y avisos.

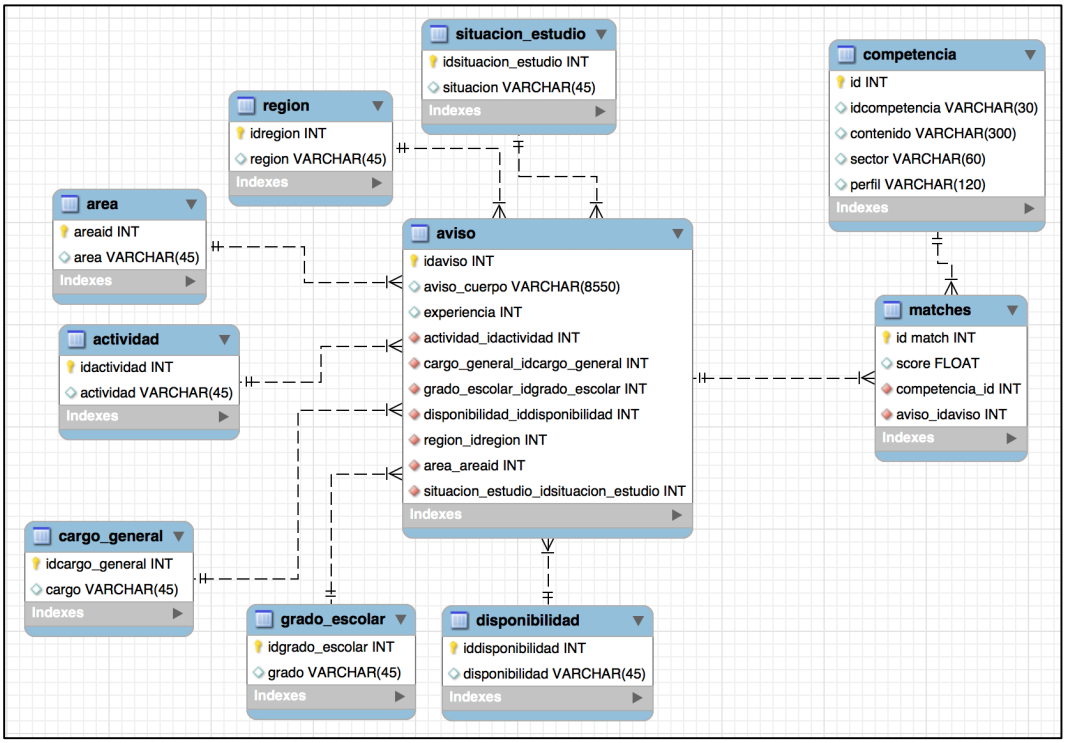


Ilustración 4.4: Modelo entidad-relación de DSA.
Fuente: Elaboración propia.

4.6. Interpretación y validación del puntaje.

Considerando la explicación y formula de cómo se obtienen los puntajes de las coincidencias se puede concluir que es un índice que cuantifica la medida en que coinciden los términos de las consultas realizadas con el contenido del aviso laboral, y que no es una relación directa u obvia con el objetivo del detector de competencias, no existe una implicancia directa de que un puntaje o umbral determinado significa que la competencia es requerida con cierto grado de certeza.

Existen formas de presentar los puntajes de forma normalizada, con un formato de porcentaje de coincidencia, pero incluso esa representación no tiene relación directa con el hecho de que si la coincidencia es efectivamente apropiada.

Lo anterior debido a que hay consideraciones respecto del contexto de la colección y la consulta, además de la intención del usuario que realiza la consulta, que pueden plasmarse o no correctamente en la sintaxis de la consulta. Estos factores impiden hacer la implicancia directa.

Entonces es necesario implementar una forma de interpretación a los puntajes obtenidos que indique cuan preciso o efectivo es el hecho de que la competencia laboral es requerida por el aviso de trabajo.

Para llevar esta tarea a cabo esto se diseñó una validación de los puntajes obtenidos asociados a las coincidencias de competencias laborales y avisos de trabajo.

En primer lugar, se diseñó la forma en que serían evaluados los puntajes, en la Ilustración 4.5 se muestra la distribución de puntajes a través de un histograma, de este se destaca la mayor concentración de puntajes en los rangos de 0 a 2 y de 2 a 4, por lo que se decidió separar el espectro de puntajes en 5 grupos, descritos en la Tabla 4.2 y representados en la Ilustración 4.6.

Grupo	Rango puntajes
1	[0, 4)
2	[4, 8)
3	[8, 12)
4	[12, 16)
5	[16, ∞]

Tabla 4.2: Rango de puntajes.
Fuente: Elaboración propia.

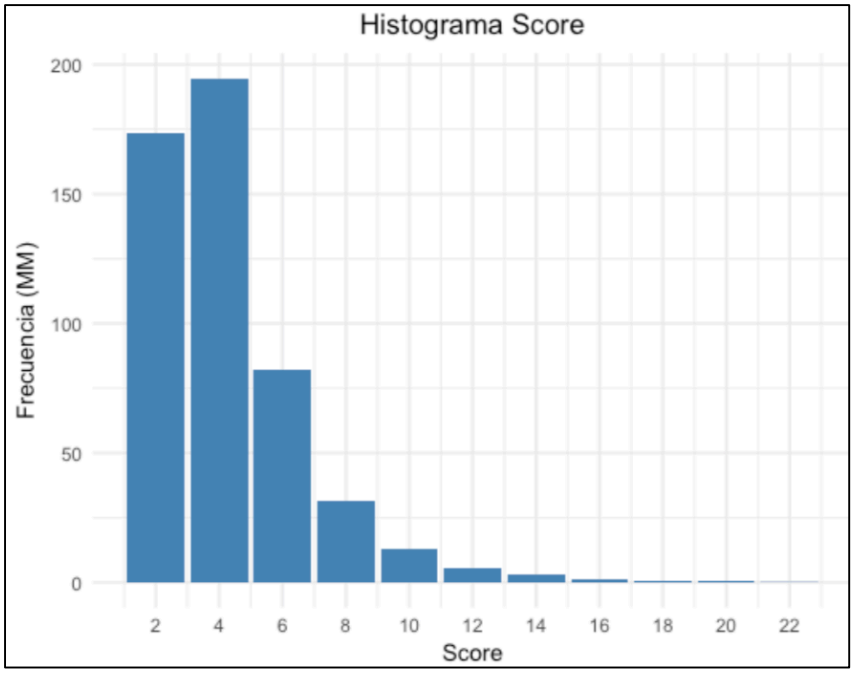


Ilustración 4.5:Histograma de puntajes de las coincidencias.
Fuente: Elaboración propia.

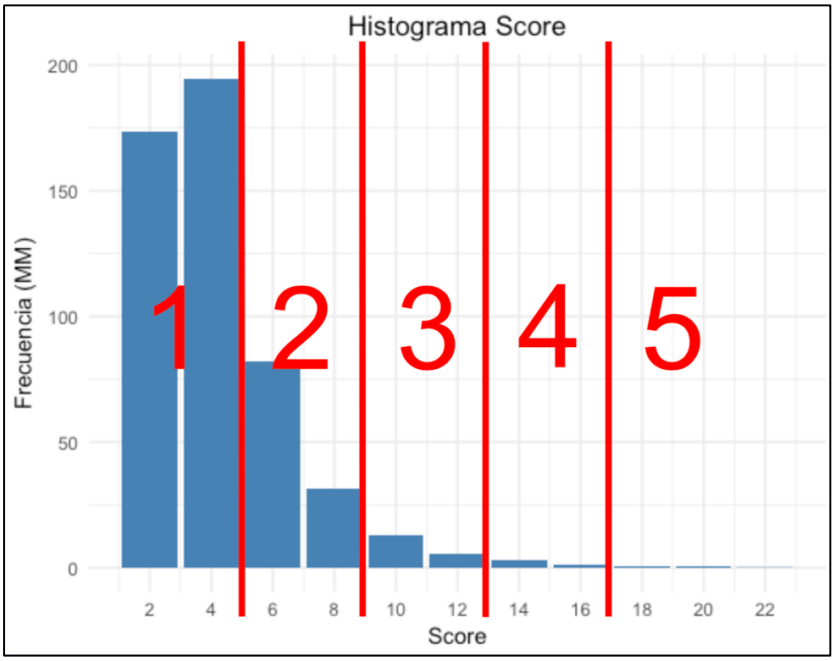


Ilustración 4.6: Representación de grupo de rangos de puntajes.
Fuente: Elaboración propia.

Luego para medir la correlación entre el rango de puntaje y la presencia de las competencias laborales en los avisos de trabajo se diseñó la siguiente clasificación:

- Al sujeto se le indica una competencia laboral.
- Se le presenta la descripción de un aviso laboral donde debiese ser requerida la competencia laboral anterior.
- Se le pregunta al sujeto si considera que la competencia laboral es requerida por el aviso de trabajo presentado.
- Este responde de forma binaria con “sí” o “no” a la pregunta.

Dado los rangos de puntajes se confeccionó una colección o repositorio de coincidencias entre competencias laborales y avisos a evaluar, esta corresponde a 20 coincidencias escogidas al azar para cada grupo, dando un total de 100 coincidencias a clasificar.

Además, se definió que 5 personas capacitadas en competencias laborales debían clasificar este repositorio, completando un total de 500 respuestas o clasificaciones que permiten evaluar el desempeño de la herramienta de detección de competencias laborales dependiendo del rango de puntajes obtenidos. Se desarrolló un sitio web para que los participantes clasifiquen las coincidencias sin necesidad de contestar presencialmente, en la Ilustración 4.6 se muestra un ejemplo de la interfaz que ellos enfrentaron para clasificar.

Los participantes fueron estudiantes universitarios de la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile, quienes recibieron una remuneración atractiva para las horas que debían invertir en la tarea.

Para garantizar en que el esfuerzo que debían emplear en la lectura y reflexión de sus respuestas sea el adecuado fueron instruidos en que sus respuestas podían ser completamente honestas además de configurar la plataforma para no permitirles responder una cierta cantidad de respuestas en un mismo día y con una separación temporal mínima entre sus respuestas.

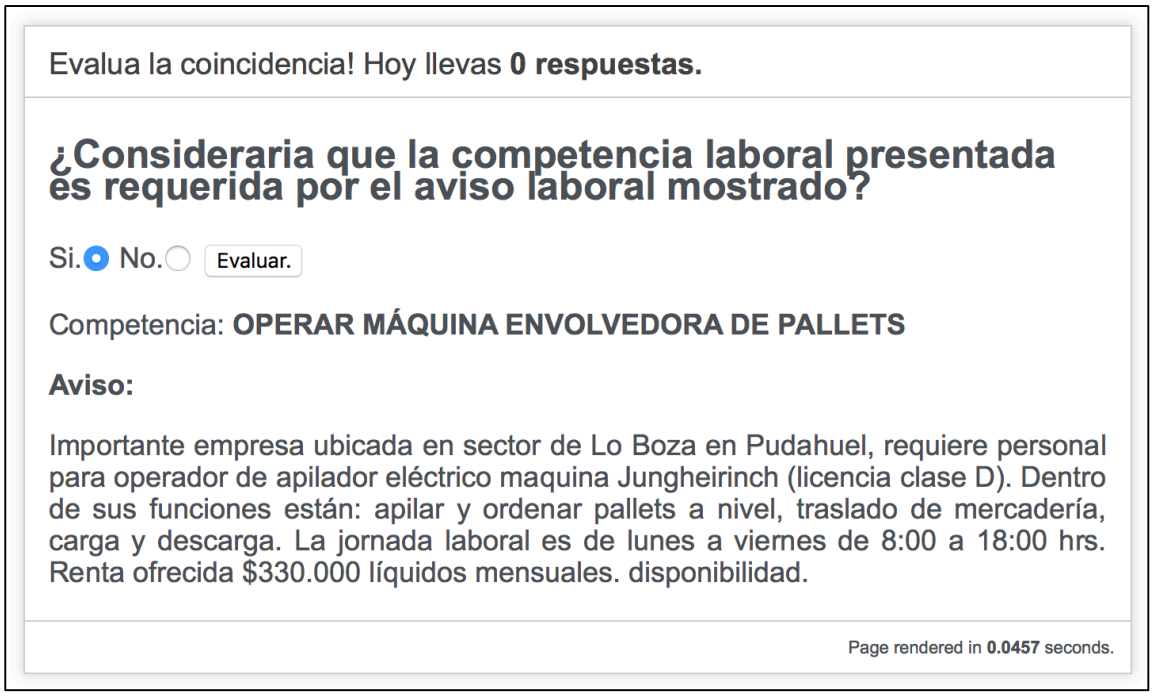


Ilustración 4.7: Interfaz de clasificación.
Fuente: Elaboración propia.

De esta forma se obtuvo para cada coincidencia del repositorio 5 evaluaciones, si 3 de estas eran respondidas positivamente marcaba la coincidencia como correctamente emparejada. En caso contrario la coincidencia encontrada por el sistema se declara no valida.

Finalmente, esta metodología permitió medir el desempeño de la herramienta de detección de competencias laborales en avisos de trabajo. Su desempeño se basa en la capacidad del sistema de determinar si cierta competencia laboral es requerida por los avisos de trabajo y con qué precisión lo ejecuta.

Cabe destacar que esta metodología fue validada con un experto en investigación de mercado, confección y análisis de encuestas. Nicolás Fritis, este experto es Profesor de la Universidad de Chile, imparte cursos de investigación de mercado y gestión de marca, además de ser director de una línea de productos a nivel nacional de la empresa IPSOS, una empresa internacional de investigación de mercado con más de 17 años de experiencia en el rubro.

Además de esta implementación, existen otras formas de medir el desempeño del sistema, una de ellas consiste en no considerar el valor numérico de cada coincidencia para su evaluación

sino tomar en cuenta el percentil al cual pertenece dentro de las coincidencias de una misma competencia laboral.

De esta forma se evitan las diferencias que conllevan las distribuciones de puntajes de cada competencia laboral.

Capítulo 5

Diseño e implementación del Datawarehouse.

En el presente capítulo se dará a entender los distintos aspectos del sistema data warehouse utilizado, desde el diseño de sus componentes hasta su implementación en un servidor accesible desde la web.

5.1. Modelo Estrella.

El modelo estrella corresponde a la estructura que tienen los datos dentro del data warehouse, en términos generales se compone de una entidad central o de hechos, conocido en inglés como "Fact Table" donde se almacenan los datos que son el objetivo principal de análisis. Luego se rodea dicha entidad de las llamadas dimensiones, estas son entidades que agregan características relacionadas a la tabla de hechos, y permiten un análisis agregado o segmentado por dichas características.

El diseño del modelo estrella a implementar es el siguiente:

- Tabla Fact o de Hechos: Corresponde a las coincidencias encontradas, es decir, la cantidad de competencias laborales detectadas en cada aviso de trabajo. En esta entidad se incluyen las llaves foráneas que hacen referencia a las dimensiones del modelo y además su variable principal para el sistema que es el puntaje obtenido en dicha coincidencia.

- Dimensiones:

    - Competencia: En esta identidad se incluye la descripción de la competencia laboral, además de información de su jerarquía, esto es el perfil laboral a cuál pertenece la competencia y el sector industrial que incluye a dicho perfil.

    - Aviso: Esta identidad corresponde a las características del aviso laboral involucrado en la coincidencia, incluye el área de la empresa a cuál está

destinado el puesto de trabajo, el tipo de cargo general que ocupará el empleado, el grado de escolaridad requerido por la empresa, el tipo de disponibilidad que exige la empresa, en cuanto a jornada full time, part time, entre otros, la actividad de la empresa que publicó el aviso y la situación actual de los estudios requerido a los postulantes.

- Región: Es la región del país desde a cuál pertenece la empresa que publica el aviso laboral involucrado en la coincidencia.

- Grupo: Esta identidad indica el rango del puntaje obtenido por la coincidencia.

- Fecha: Esta identidad incluye la fecha de publicación del aviso laboral involucrado, almacenando su mes, año, semana, trimestre en cual fue publicado.

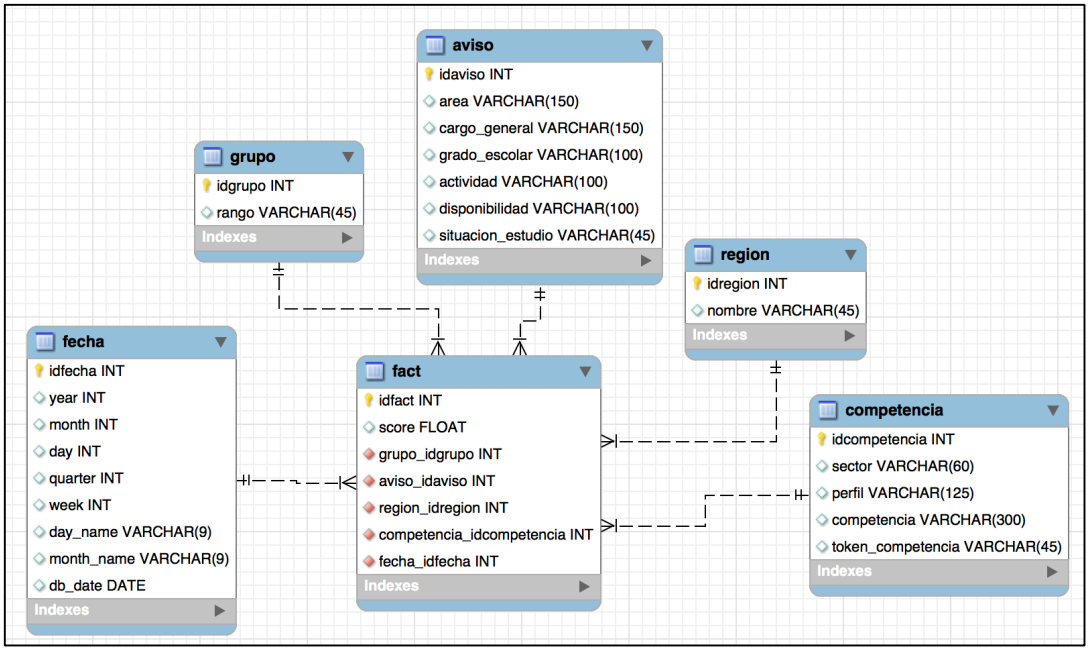


Ilustración 5.1: Modelo estrella del Data warehouse.
Fuente: Elaboración propia.

Su implementación se realizó en el mismo sistema en el que fue implementado el Data Staging Area (DSA), un sistema de administración de base de datos basado en MariaDB.

A continuación, se realiza una descripción de los elementos de cada entidad o dimensión del modelo estrella:

- Grupo: 5 registros con los rangos de los puntajes.

- Región: 19 registros correspondientes a las regiones territoriales chilenas, y algunas extranjeras.

- Fecha: 2.922 registros de fechas que corresponden a las fechas de publicación de los avisos laborales, además se segmenta la información por año, mes, día, semana y trimestre del año.

- Competencia: 2.150 registros de competencias laborales, donde se especifica el sector industrial a cuál pertenece dicha competencia, el perfil profesional relacionado, la descripción de la competencia y finalmente el identificador asignado en el catálogo generado por ChileValora.

- Avisos: 685.219 registros de los avisos laborales publicados en la Bolsa Nacional de Empleo utilizados para la detección de competencias laborales. Los atributos de esta entidad corresponden a el grado escolar requerido, el área del cargo ofrecido, el tipo de cargo, la disponibilidad horaria requerido, el tipo de situación de estudio y finalmente el tipo de actividad de la empresa que publicó dicho aviso.

- Fact: Una cantidad de 394.614.824 de registros, donde se indica el puntaje asociado a la coincidencia y las llaves foráneas que relacionan la coincidencia con las dimensiones anteriormente descritas.

5.2. Extracción, Transformación y Carga de los datos.

Para poblar el modelo estrella, se debió crear y ejecutar un proceso ETL, para cada entidad de este modelo. Gracias al modelo de datos del Data Staging Area, que se puede encontrar en la Ilustración 4.4, permitió una extracción y carga directa desde el DSA a las siguientes dimensiones:

- Fecha.

- Competencia.

- Grupo.

- Región.

Las entidades Avisos y Fact necesitan de una metodología distinta. Para la entidad Avisos se usó un programa propio escrito en Java que se conecta con ambas bases de datos, DSA y el modelo estrella, luego por cada registro de la entidad del DSA, se ejecuta una consulta que cruzaba dicha entidad con las que la rodean, emparejando a través de la llave foránea. De esta

forma se remplazan de los atributos las llaves foráneas por su atributo especifico, es decir, si la llave foránea era “5” que indicaba al registro “Profesor” solo se almacena el atributo “Profesor”.

Se realizó un proceso similar para la entidad Fact, por cada registro dentro de la identidad Matches, del DSA, mediante consultas SQL y un programa Java, se seleccionó su identificador, su puntaje asociado, además de la competencia laboral y el aviso involucrado en la coincidencia. Utilizando el identificador del aviso se seleccionó las llaves foráneas correspondientes a las dimensiones del modelo estrella. Finalmente se realizó una inserción al modelo estrella.

5.3. Plataforma Business Intelligence.

Como se explicó en el Capítulo 3.4, la plataforma de BI escogida para la implementación de la plataforma de análisis fue la versión comunitaria de Saiku, debido principalmente a su interfaz y características amigables al usuario, además de su API que permite una comunicación entre aplicaciones o servicios computacionales con el servidor para ejecutar consultas al sistema o generar reportes.

Para implementar dicha arquitectura es necesario configurar dos componentes, por un lado, la conexión a la base de datos y por otro el esquema que define el modelo lógico detrás de la arquitectura.

5.3.1. Conexión a la base de datos.

Para conectar el sistema data warehouse con la fuente de datos es necesario especificar la siguiente configuración:

- Tipo de conexión: Mondrian.
- URL: jdbc:mariadb://localhost:3306/modelo_estrella. En este campo se indica la URL que utiliza el sistema para conectarse a la base de datos.
- Esquema: DW. Acá se especifica que esquema cargado en el servidor utilizara esta conexión.
- JDBC.Driver: org.mariadb.jdbc.Driver. Especifica el driver que utiliza el administrador de base de datos para manejar las conexiones y consultas a este mismo.
- Usuario: El usuario a utilizar para la conexión, debe tener privilegios, al menos, de lectura de la base de datos.

- Contraseña: Corresponde a la contraseña del usuario para la conexión a la base de datos.

Esta configuración habilita al servidor data warehouse conectarse a la fuente de datos y hacer consultas a estos, variara dependiendo de en qué sistema fue implementada la base de datos.

5.3.2. Esquema Mondrian.

Como fue descrito en el Capítulo 3.5, es necesario configurar el modelo lógico de los datos, esto corresponde a sus métricas, sus dimensiones, las relaciones entre estas y el modelo físico de donde se encuentra la información.

Esta configuración se traduce a un archivo XML, que es explicado a continuación, se desglosara segmentado por sus secciones con la finalidad de comprender la finalidad de cada sección en la implementación del modelo estrella en el sistema data warehouse.

Para definir esquema físico se realizó la configuración de la Ilustración 5.2. En esta se especifica cada entidad con su nombre exacto correspondiente al que posee en el modelo estrella.

```
<PhysicalSchema>
    <Table name="competencia" />
    <Table name="aviso" />
    <Table name="fact" />
    <Table name="grupo" />
    <Table name="region" />
    <Table name="fecha" />
</PhysicalSchema>
```

Ilustración 5.2:Componente del Esquema físico del esquema Mondrian.
Fuente: Elaboración propia.

Luego se procede a definir el cubo a utilizar en el sistema, dentro de este se configuran las dimensiones del modelo estrella. En la Ilustración 5.3 se ejemplifica con solo una dimensión, en este caso “Competencia”. Es necesario mediante atributos configurar la tabla del esquema físico que contiene la información de esta dimensión, el atributo que corresponde a su llave primaria y luego los atributos que sirven para diseccionar el análisis indicando su atributo en la entidad.

```
<Dimension name="Competencia" table="competencia" key="CompetenciaId">
        <Attributes>
                <Attribute name="Sector" column='sector' hierarchyHasAll="true"  >
                        <Key><Column name="sector" /></Key>
                </Attribute>
                <Attribute name="Contenido" column="competencia" >
                        <Key><Column name="competencia" /></Key>
                </Attribute>
                <Attribute name="CompetenciaId" column='idcompetencia' >
                        <Key><Column name="idcompetencia" /></Key>
                </Attribute>
        </Attributes>
</Dimension>
```

Ilustración 5.3: Componente dimensión en el esquema Mondrian.
Fuente: Elaboración propia.

Otra componente que es necesaria configurar, son las métricas que el sistema data warehouse medirá y será analizada por el usuario, esta se presenta en la Ilustración 5.4. En esta componente se indica la columna que será transformada o se le aplicará una función para dar como resultado la métrica deseada. En el atributo "aggregator" se da la instrucción de qué tipo de función se aplica en la métrica, esta puede ser una suma simple, un conteo, un promedio, entre otros.

```
<Measures>
    <Measure name="Cantidad de Competencias Laborales" column="idfact" aggregator = "count" />
    <Measure name="Cantidad Avisos" column="aviso_idaviso" aggregator = "distinct-count" />
</Measures>
```

Ilustración 5.4: Componente de métricas en el esquema Mondrian.
Fuente: Elaboración propia.

Finalmente es necesario relacionar las dimensiones con la tabla de hechos y así permitir que las métricas definidas sean estudiadas a lo largo de estas. Para ellos se definen estas relaciones mediante enlaces, donde se explicita el atributo de la tabla de hechos que corresponde a la llave foránea de cada dimensión y además se le asigna un nombre que luego será el que observará el usuario. Esta componente se puede observar en la Ilustración 5.5.

```
<DimensionLinks>
    <ForeignKeyLink dimension="Competencia" foreignKeyColumn="competencia_idcompetencia"/>
    <ForeignKeyLink dimension="Puntaje" foreignKeyColumn="grupo_idgrupo"/>
    <ForeignKeyLink dimension="Region" foreignKeyColumn="region_idregion"/>
    <ForeignKeyLink dimension="Avisos" foreignKeyColumn="aviso_idaviso"/>
    <ForeignKeyLink dimension="Fecha" foreignKeyColumn="fecha_idfecha"/>
</DimensionLinks>
```

Ilustración 5.5: Componente de relaciones de dimensiones.
Fuente: Elaboración propia.

En conjunto todas estas componentes conforman el esquema Mondrian utilizado para el sistema data warehouse, la completitud de este esquema se encuentra en el Anexo B.

5.3.3. Interfaz de usuario.

La configuración anteriormente descrita da como resultado una plataforma de inteligencia de negocios o B.I. (por sus siglas en ingles), accesible para el usuario mediante un navegador web.

En un principio debe ingresar con un nombre de usuario y contraseña configurados previamente por el administrador del servidor. Luego el usuario debe escoger qué Cubo desea explorar, en el caso de esta memoria el cubo fue nombrado “Competencias Laborales v2”.

Una vez escogido el Cubo a explorar o analizar al usuario se le muestra una interfaz, parte de ella está incluida en la Ilustración 5.6, a continuación, se explicará las componentes más relevantes en esta interfaz que corresponden a los indicadores de la misma Ilustración 5.6.

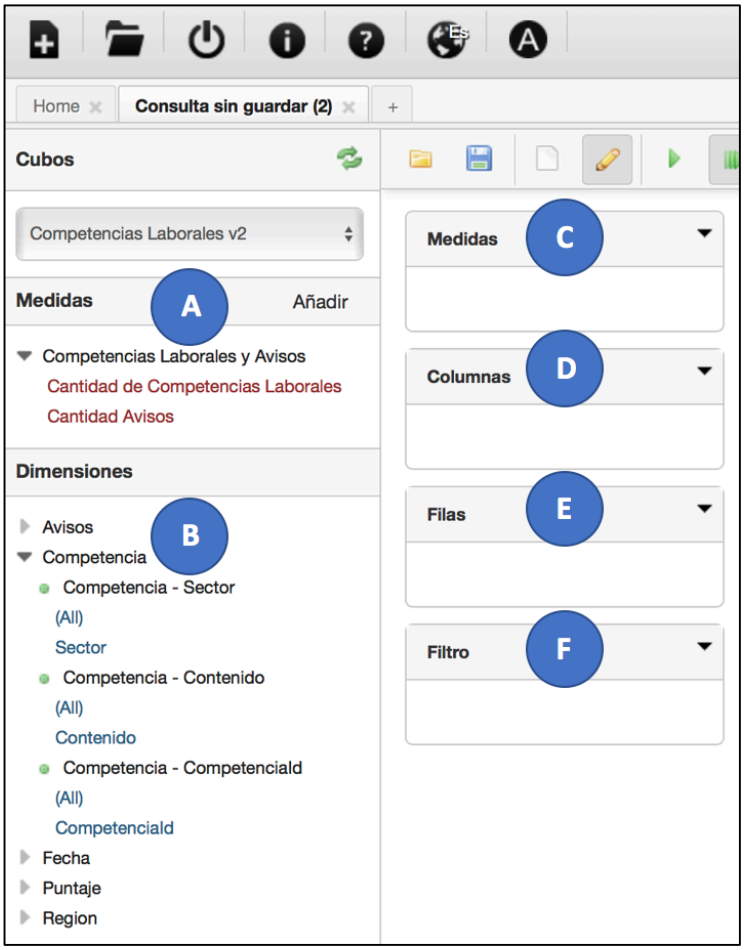


Ilustración 5.6: Interfaz inicial plataforma BI.
Fuente: Elaboración propia.

A. Medidas: En esta sección se escogen las medidas a calcular, la principal del sistema es "Cantidad de Competencias Laborales" y en segundo lugar "Cantidad de Avisos", estas corresponden a las métricas definidas en el esquema Mondrian.

B. Dimensiones: Esta sección permite al usuario escoger las dimensiones que segmentarán las medidas o métricas, como se muestra en la misma Ilustración 5.6, la dimensión "Competencia" o cualquier otra puede ser seleccionada y se despliegan los atributos de ésta. Estas pueden ser arrastradas hacia las componentes Columnas, Filas y/o Filtro.

C. Medidas: Esta sección difiere del punto A, aquí aparecerán las métricas que fueron seleccionadas y pueden ser removidas arrastrándolas afuera de su espacio o cliqueándolas nuevamente.

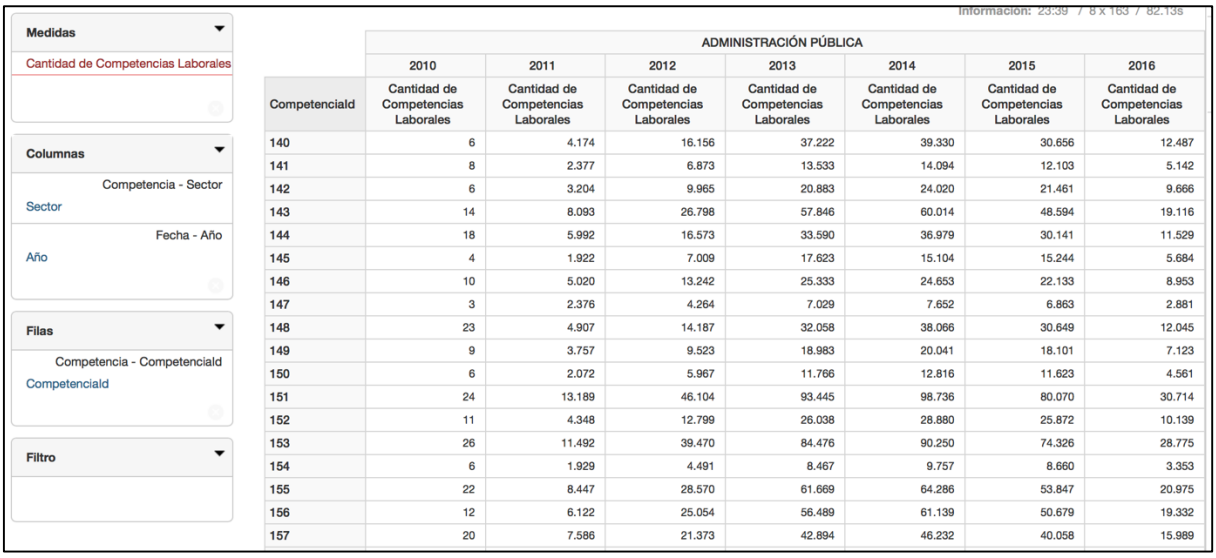
D. Columnas: Aquí se agregan las dimensiones que segmentarán la información en columnas para su análisis. Estas pueden ser arrastradas dentro de la sección o fuera por el usuario.

E. Filas: En esta sección se añaden las dimensiones para segmentar nuevamente el análisis en filas.

F. Filtro: Este módulo permite al usuario agregar dimensiones para luego seleccionarlas y encontrar un nuevo módulo para seleccionar que miembros del atributo seleccionado de la dimensión serán incluidos o excluidos del análisis.

Las componentes mínimas que se deben configurar o seleccionar por un usuario para realizar una consulta son seleccionar una medida a evaluar y luego un atributo de alguna dimensión. Además, las componentes descritas anteriormente permiten al usuario realizar distintas consultas al sistema, dependiendo de qué aspectos quiere evaluar, sin esta interfaz amigable el usuario debiese tener conocimientos en consultas MDX para realizar cualquier tipo de análisis. De todas formas, se presenta un recurso desarrollado por Microsoft que explica como confeccionar consultas MDX [47]

Un ejemplo de análisis y uso de la plataforma BI se puede encontrar en la Ilustración 5.7. En ella se ejecuta una consulta por la cantidad de competencias laborales demandadas, en el sector "Administración pública", segmentado por año y para cada competencia laboral.



	ADMINISTRACIÓN PÚBLICA						
	2010	2011	2012	2013	2014	2015	2016
CompetenciaId	Cantidad de Competencias Laborales	Cantidad de Competencias Laborales	Cantidad de Competencias Laborales	Cantidad de Competencias Laborales	Cantidad de Competencias Laborales	Cantidad de Competencias Laborales	Cantidad de Competencias Laborales
140	6	4.174	16.156	37.222	39.330	30.656	12.487
141	8	2.377	6.873	13.533	14.094	12.103	5.142
142	6	3.204	9.965	20.883	24.020	21.461	9.666
143	14	8.093	26.798	57.846	60.014	48.594	19.116
144	18	5.992	16.573	33.590	36.979	30.141	11.529
145	4	1.922	7.009	17.623	15.104	15.244	5.684
146	10	5.020	13.242	25.333	24.653	22.133	8.953
147	3	2.376	4.264	7.029	7.652	6.863	2.881
148	23	4.907	14.187	32.058	38.066	30.649	12.045
149	9	3.757	9.523	18.983	20.041	18.101	7.123
150	6	2.072	5.967	11.766	12.816	11.623	4.561
151	24	13.189	46.104	93.445	98.736	80.070	30.714
152	11	4.348	12.799	26.038	28.880	25.872	10.139
153	26	11.492	39.470	84.476	90.250	74.326	28.775
154	6	1.929	4.491	8.467	9.757	8.660	3.353
155	22	8.447	28.570	61.669	64.286	53.847	20.975
156	12	6.122	25.054	56.489	61.139	50.679	19.332
157	20	7.586	21.373	42.894	46.232	40.058	15.989

Ilustración 5.7: Ejemplo de consulta a la plataforma BI.
Fuente: Elaboración propia.

La plataforma, además de permitir al usuario ejecutar consultas al sistema, tiene una serie de utilidades con objetivo de facilitar la tarea de entregar reportes acerca de la información analizada. Luego de la ilustración 5.8 se encuentra una descripción de estas herramientas.

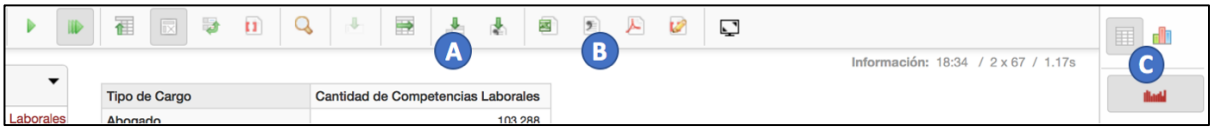


Ilustración 5.8: Utilidades de la plataforma.
Fuente: Elaboración propia.

A. Estas herramientas permiten al usuario hacer un análisis detallado de alguna celda o dato en particular.

B. En esta sección el usuario puede exportar los resultados de la consulta realizada en formato PDF, Excel y CSV (archivo separado por comas).

C. Esta herramienta es de gran utilidad, permite cambiar la vista de los resultados entre Tablas con información a una serie de distintos gráficos, los cuales se pueden exportar en formato PNG, JPEG o PDF.

5.4. Indicadores de la plataforma BI.

Si bien la plataforma desarrollada permite al usuario realizar una gran cantidad de análisis, segmentando temporalmente, por competencia, sector económico, tipo de cargo, entre otros, también es capaz de incluir varias dimensiones simultáneamente.

Se proponen una serie de indicadores a partir de las métricas definidas que permiten evaluar distintas situaciones. Con objetivo de mostrar ejemplo de cómo puede ser utilizado la plataforma de análisis para apoyar a la gestión.

Evolución de demanda de competencias laborales por año.

Este indicador muestra la cantidad demandada en el sistema por competencia laboral en cada año, permite saber si está siendo más vigente o más obsoleta, lo que puede apoyar decisiones sobre si fomentar o no las capacitaciones en esas competencias laborales.

Inicialmente se incluye como dimensiones, los años a analizar (por defecto desde 2011 hasta 2016), luego el identificador de la competencia laboral para diferenciar entre ellas, luego como es necesario agregar como filtro el grupo de puntajes de la coincidencia y dejar solo el grupo 5. Adicionalmente se pueden agregar como filtros el tipo de cargo, el área industrial, la actividad de la empresa, entre otros, para hacer un análisis más específico.

El indicador explicado se traduce en la consulta MDX de la Ilustración 5.9, gracias a la plataforma BI el usuario no requiere de escribir manualmente dicha consulta para calcular el indicador mostrado. Si no que la plataforma la genera por él, mientras hace selección de las dimensiones, filtros y métricas a utilizar.

```
WITH
SET [~FILTER] AS
  {[Puntaje].[grupoID].[5]}
SET [~COLUMNS] AS
  {[Fecha].[Año].[Año].Members}
SET [~ROWS] AS
  {[Competencia].[CompetenciaId].[CompetenciaId].Members}
SELECT
NON EMPTY CrossJoin([~COLUMNS], {[Measures].[Cantidad de Competencias Laborales]}) ON COLUMNS,
NON EMPTY [~ROWS] ON ROWS
FROM [Competencias Laborales v2]
WHERE [~FILTER]
```

Ilustración 5.9: Consulta MDX de la evolución de demanda de competencias laborales. Fuente: Elaboración propia.

Un ejemplo de este indicador se muestra en la Ilustración 5.10, este no incluye la totalidad de los datos generados en el sistema.

	2010	2011	2012	2013	2014	2015	2016
CompetenciaId	Cantidad de Competencias Laborales	Cantidad de Competencias Laborales	Cantidad de Competencias Laborales	Cantidad de Competencias Laborales	Cantidad de Competencias Laborales	Cantidad de Competencias Laborales	Cantidad de Competencias Laborales
1		38	76	143	123	92	28
2		20	23	97	33	38	14
3		36	56	123	105	96	37
4		16	37	159	258	114	53
5		7	21	35	22	10	6
6	1	22	84	174	145	117	48
7	1	196	527	1,018	957	879	311
8		13	29	60	86	76	26
9		6	4	10	7	6	2
10		2	1	7	5	15	14
11		10	24	7	15	10	3
12		22	121	306	424	423	115
16	1	21	82	155	131	94	44
17			7	4	15	8	4
18		17	43	85	113	94	33
20		11	61	81	50	97	21
21		31	128	219	254	291	77
22		20	50	159	255	124	48
23		41	114	303	294	205	56
26		108	363	623	703	897	531
27	1	78	251	480	794	788	335

Ilustración 5.10: Ejemplo de resultado del indicador de evolución de demanda por competencia laboral. Fuente: Elaboración propia.

Demanda de competencias laborales por tipo de cargo.

El siguiente indicador muestra la demanda de competencias laborales segmentando por tipo de cargo general del aviso de trabajo. Básicamente, muestra si las competencias laborales son demandadas para distintos tipos de cargos, sean administrativo, gerenciales o si son de un perfil más cercano a trabajos operarios o de servicio.

Este indicador puede servir a los impulsores de programas de capacitación para saber en qué tipo de perfil laboral enfocar sus programas, donde ofrecerlos y que elementos o enfoque deben poseer.

Este indicador se traduce en la consulta MDX de la Ilustración 5.11 y un ejemplo de cómo se visualiza este indicador en la plataforma de análisis se puede observar en la Ilustración 5.12.

```
WITH
SET [~FILTER] AS
  {[Puntaje].[grupoID].[5]}
SET [~COLUMNS] AS
  {[Avisos].[Tipo de Cargo].[Tipo de Cargo].Members}
SET [~ROWS] AS
  {[Competencia].[CompetenciaId].[CompetenciaId].Members}
SELECT
NON EMPTY CrossJoin([~COLUMNS], {[Measures].[Cantidad de Competencias Laborales]}) ON COLUMNS,
NON EMPTY [~ROWS] ON ROWS
FROM [Competencias Laborales v2]
WHERE [~FILTER]
```

Ilustración 5.11: Consulta MDX de la demanda de competencias laborales por tipo de cargo. Fuente: Elaboración propia.

Nuevamente esta consulta MDX no debe ser escrita por el usuario, es más, no necesita entenderla para hacer uso de la plataforma de análisis y obtener los resultados o análisis deseados. Esto demuestra las ventajas de implementar una plataforma BI para el análisis de demanda de competencias laborales, habilitando a una gran cantidad de tipos de usuario no calificados en estas tecnologías para usar una herramienta de apoyo a la toma de decisión.

	Abogado	Abogado Independiente	Administrador	Administrativo	Agente	Ama de casa	Analista	Asesor	
CompetenciaId	Cantidad de Competencias Laborales	Cantidad de Competencias Laborales	Cantidad de Competencias Laborales	Cantidad de Competencias Laborales	Cantidad de Competencias Laborales	Cantidad de Competencias Laborales	Cantidad de Competencias Laborales	Cantidad de Competencias Laborales	C
1			3	11	1		30	4	
2			4	35	2		19	2	
3			9	75	1		56	17	
4			6	23			2	5	
5			3	12			41		
6	1		20	31			15	9	
7	4		109	1,515	2		170	6	
8			3	61			6	6	
9								2	
10							1		

Ilustración 5.12: Ejemplo de resultado del indicador de demanda de competencias laborales por tipo de cargo. Fuente: Elaboración propia.

Estos dos indicadores son un ejemplo de cómo el sistema de análisis de demanda de competencias laborales desarrollado puede ser utilidad en el apoyo a la toma de decisiones de las entidades encargadas de fomentar y financiar distintos programas de capacitaciones.

Capítulo 6

Resultados.

Este capítulo se encarga de numerar, reportar y describir los resultados obtenidos producto del trabajo de esta memoria, tanto los entregables finales como los resultados obtenidos por el mismo sistema de detección y análisis de competencias laborales.

Los resultados presentados a continuación se dividen en el motor de búsqueda SOLR o sistema de detección de competencias laborales, la plataforma de análisis de competencias laborales o plataforma B.I. y finalmente el desempeño y validación de los puntajes asociados a las coincidencias entre competencias laborales y avisos de trabajo.

6.1. Sistema de detección de competencias laborales.

En cuanto al sistema de detección de competencias laborales se reporta como resultado el servidor de búsqueda configurado, junto a las consultas, para detectar competencias laborales en avisos de trabajos. Junto con el servidor o motor de búsqueda se reporta el programa encargado de seleccionar las competencias laborales, ejecutar las consultas al motor SOLR, recuperar las coincidencias encontradas y almacenarlas en el DSA.

Además, el motor de búsqueda permite consultas a través de una interfaz para el usuario, accesible desde un navegador web, o mediante comunicación http con alguna otra aplicación o programa. Aplicando una serie de transformaciones y técnicas de textmining para una búsqueda eficiente.

La interfaz inicial del motor de búsqueda y la de consultas se pueden observar en las Ilustraciones 6.1 y 6.2. En la primera ilustración se puede observar en el atributo “Num Docs”, número de documentos, que indica que el sistema posee 684.639 avisos laborales como parte de su repositorio para la búsqueda.

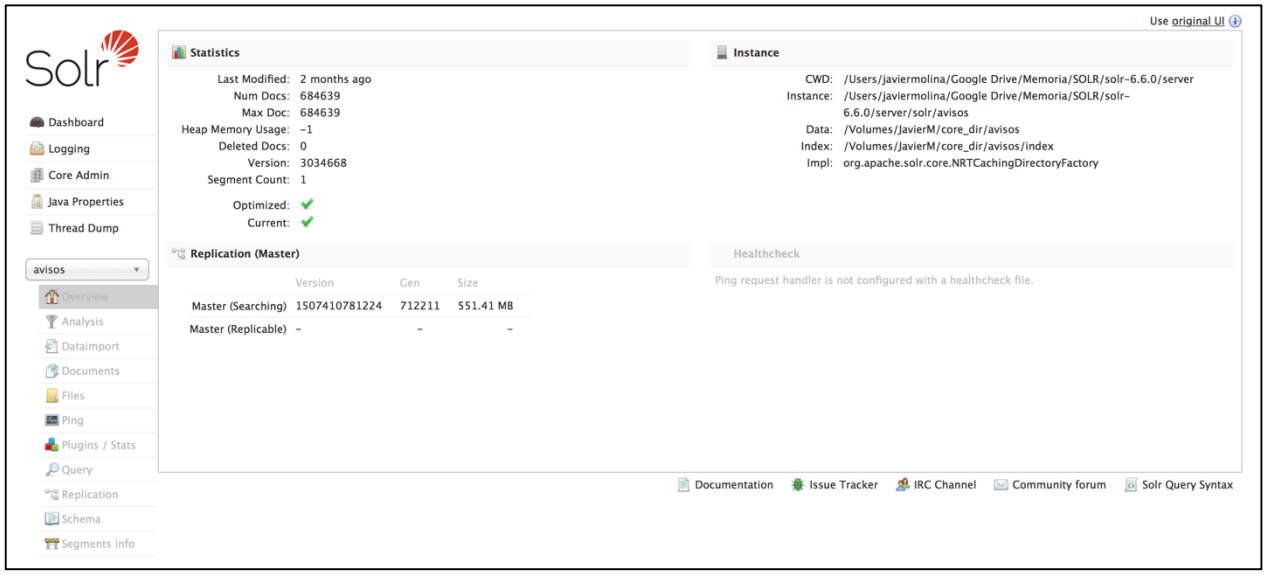


Ilustración 6.1: Interfaz inicial motor SOLR.
Fuente: Elaboración propia.

En la segunda ilustración (Ilustración 6.2), se incluye un ejemplo de consulta al sistema, donde se buscan las coincidencias de la competencia laboral "Administrar los procedimientos de higiene, seguridad y emergencia." Esta produce cerca de 137.000 coincidencias en el repositorio de avisos laborales. Este es un ejemplo del ejercicio que reprodujo el programa principal de detección de competencias para cada una de las competencias laborales del sistema.

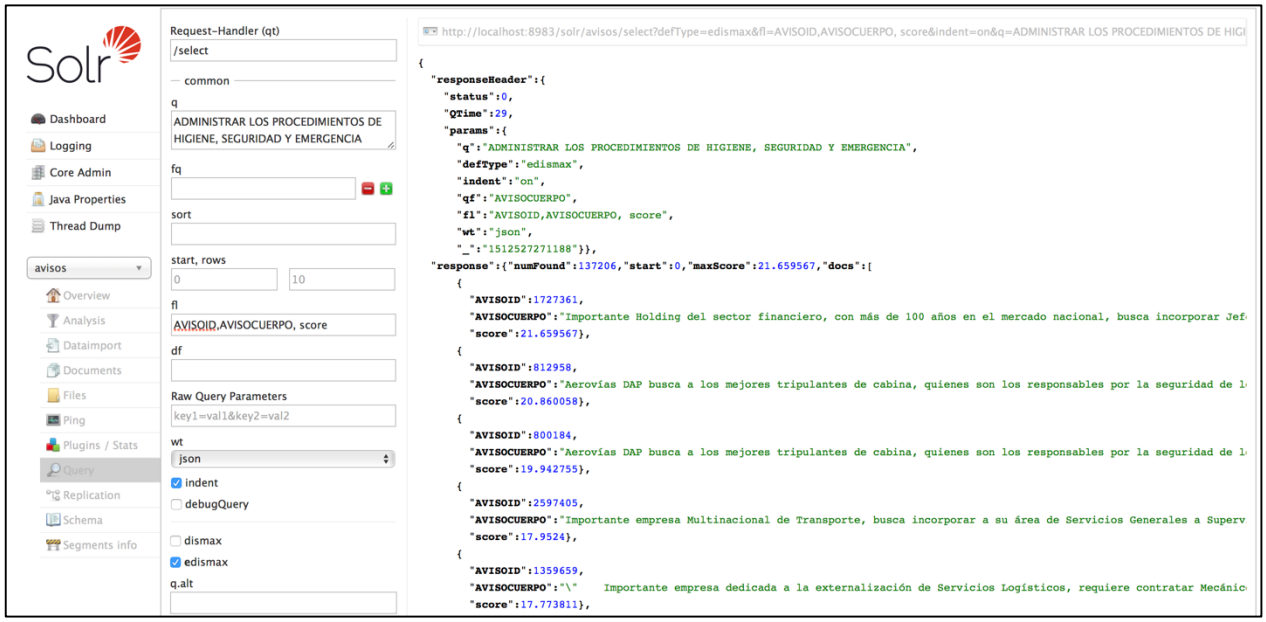


Ilustración 6.2: Interfaz de consulta del motor SOLR.
Fuente: Elaboración propia.

Finalmente se reportan los resultados de la aplicación del sistema de detección sobre la base de datos de avisos laborales de la BNE, Bolsa Nacional de Empleo, y los repositorios confeccionados de avisos laborales y competencias laborales. Estos se encuentran en la Tabla 6.1.

	Cantidad
Competencias Laborales	2.150
Avisos Laborales	684.639
Coincidencias Competencias-Avisos	394.614.824
Coincidencias por Aviso Laboral	183.542

Tabla 6.1: Estadística general.
Fuente: Elaboración propia.

A continuación, se reporta la distribución de las coincidencias en los rangos de puntajes asociados a la calidad de las coincidencias entre avisos laborales y competencias laborales, definidos en el sub capítulo 4.6. Esta distribución se reporta en la Tabla 6.2.

Grupo	Rango	Cantidad de Coincidencias	Coincidencias por Competencia	Coincidencias por Aviso Laboral
1	[0, 4)	264.760.775	123.145	387
2	[4, 8)	104.698.431	48.697	153
3	[8, 12)	19.155.381	8.909	28
4	[12, 16)	4.385.837	2.040	6
5	[16, ∞]	1.614.980	751	2

Tabla 6.2: Distribución de coincidencias por puntaje.
Fuente: Elaboración propia.

6.2. Desempeño de la detección de competencias laborales.

Esta sección da a conocer los resultados de la clasificación del desempeño de los puntajes obtenidos para las coincidencias entre competencias laborales y avisos de trabajo. Dicha clasificación constó de 5 personas capacitadas para evaluar la presencia de las competencias laborales en los avisos de trabajos en una ventana temporal de 5 días.

Estas 5 personas clasificaron como coincidencias correctas o incorrectas, 100 coincidencias generadas por el sistema de detección de competencias laborales, distribuidas equitativamente en cada rango, resultando en 20 coincidencias por rango.

Finalmente hay 5 clasificaciones por cada coincidencia, una por cada participante, si 3 de ellas son positivas se considera la coincidencia como correcta, en caso contrario se considera

incorrecta la detección. Esto genera información suficiente para medir el desempeño del sistema de detección de competencias laborales.

Los resultados de la clasificación se encuentran en la Tabla 6.3. segmentados por grupo de puntajes. Siendo 5 el grupo con mayor puntaje (de 16 o mayores puntajes) y 1 el grupo con menor puntaje (de 0 a 4).

Grupo	Positivos	Negativos	Correctos	Incorrectos
1	2	18	10 %	90 %
2	6	14	30 %	70 %
3	11	9	55 %	45 %
4	12	8	60 %	40 %
5	19	1	95 %	5 %

Tabla 6.3: Resultados de desempeño del sistema de detección de competencias laborales.
Fuente: Elaboración propia.

De los resultados se desprende una tendencia positiva en el desempeño de la clasificación en cuanto al grupo de puntajes, lo que es esperable dada la construcción del puntaje, que fue creado para medir el nivel de coincidencia como fue explicado en el sub capítulo 4.3.2.

El grupo 5 presenta un porcentaje de detección correcto del 95%, al ser comparado con la cantidad de competencias por aviso laboral que entrega este grupo (Tabla 6.2) que indica un promedio de 2 competencias laborales, resulta razonable considerar esa cantidad de competencias por aviso laboral en promedio. Comparado con el grupo 4, que son 6 competencias laborales en promedio por aviso laboral.

Se realiza además un análisis en conjunto por grupos, agrupando de mayor a menor puntaje los grupos. Esto se reporta en la Tabla 6.4.

Grupos	Positivos	Negativos	Correctos	Incorrectos
5	19	1	95,0 %	5,0 %
5+4	31	9	77,5 %	22,5 %
5+4+3	42	18	70,0 %	30,0 %
5+4+3+2	48	32	60,0 %	40,0 %
Todos	50	50	50,0 %	50,0 %

Tabla 6.4: Desempeño de grupos en conjunto.
Fuente: Elaboración propia.

De los resultados reportados en la tabla anterior se observa una concordancia entre agrupar rangos de menor puntaje con su evaluación por separado.

Si bien, el desempeño de la segunda agrupación del grupo 5 y 4, posee un buen desempeño 77,5% de detecciones correctas, éstas implican que en promedio cada aviso de trabajo tiene relacionado 8 competencias laborales en su descripción, lo que es alto considerando el largo promedio de la descripción del aviso laboral.

Finalmente, considerando el desempeño del grupo 5 es de un 95% de detección correcta, junto a la cantidad de competencias laborales por aviso laboral, 2 competencias en cada aviso en promedio, en comparación a la agrupación del grupo 4 y 5, que posee un 77,5% de detecciones correcta, y un promedio de 8 competencias laborales por cada aviso de trabajo.

Se recomienda que para todos los análisis hechos sobre el sistema desarrollado en esta memoria se utilice el grupo 5, esto incluye solo las coincidencias con un puntaje mayor o igual a 16.

6.3. Plataforma de análisis de competencias laborales.

Los resultados obtenidos de la plataforma de análisis de competencias laborales se dividen en dos elementos principales:

A. En primer lugar, está la plataforma tecnológica que permite al usuario disponer de una interfaz, accesible desde cualquier lugar físico mediante un navegador web y conexión a internet. Esta plataforma permite realizar distintos análisis a las competencias laborales detectadas en la información de la BNE.

B. El segundo elemento corresponde a la capacidad de análisis de la plataforma, en cuanto a cantidad de información que puede analizar y además por las características o dimensiones por las cuales se puede segmentar el análisis.

En cuanto al primer resultado, la interfaz del usuario hacia la plataforma, se entrega un acceso mediante un navegador web a la plataforma de análisis. Para ingresar a esta se debe dirigir a la dirección http://javiermolina.wic.cl e ingresar con un usuario y contraseña provista por la administración de la plataforma, en primera instancia, esta corresponde al personal del Web Intelligence Centre. En la Ilustración 6.3 se observa la página de inicio de la plataforma de análisis, que corresponde a la página para ingresar a la plataforma.

Luego de ingresar, el usuario ya puede hacer uso de la plataforma para sus distintos análisis y reportes, estos pueden ser extraídos como archivos Excel, CSV o PDF. Los detalles de las operaciones se encuentran en el sub capítulo 5.3.3.

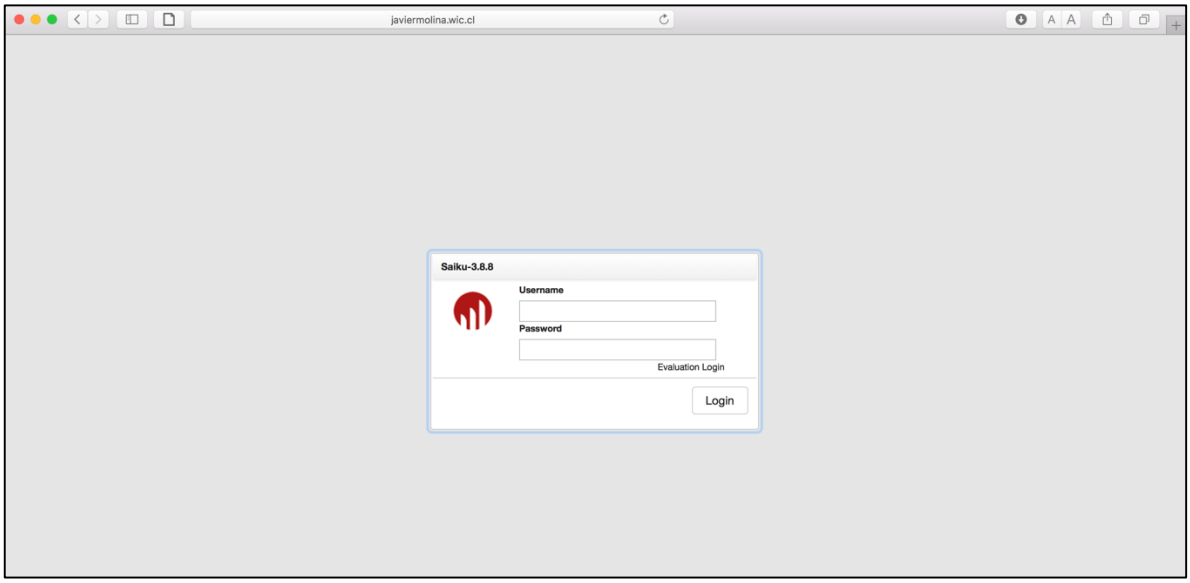


Ilustración 6.3: Vista de la página inicial de la plataforma de análisis desde un navegador web.
Fuente: Elaboración propia

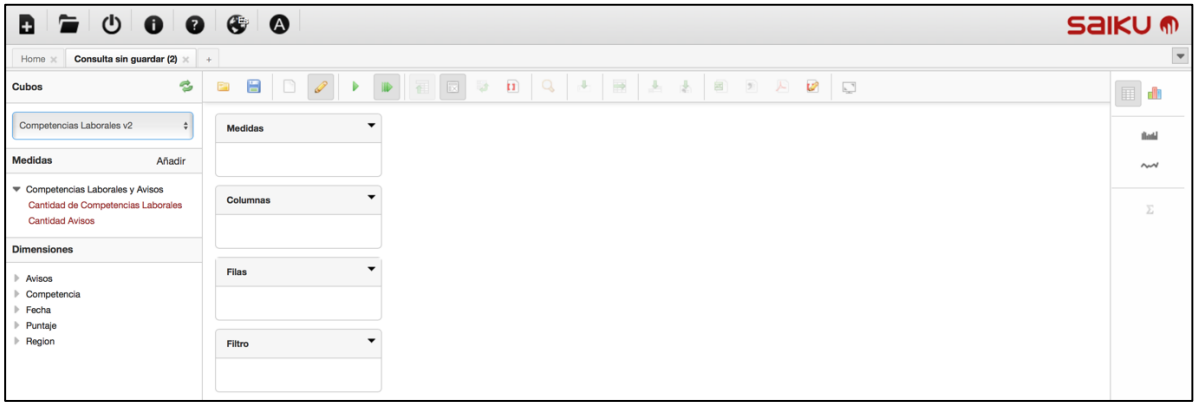


Ilustración 6.4 Interfaz de consultas al sistema desde un navegador web.
Fuente: Elaboración propia

Pasando al segundo elemento de los resultados de la plataforma de análisis de competencias laborales se reportan sus siguientes características de los análisis posibles:

- Posee 2 métricas:
    - Cantidad de Competencias Laborales.
    - Cantidad de Avisos.
- Existen 5 dimensiones por las cuales segmentar el análisis:
    - Fecha.

- Puntaje.
- Región.
- Competencias.
- Avisos.

Se debe hacer un hincapié en ciertas dimensiones, si bien hay dimensiones sencillas como lo son Puntaje, Fecha y Región permiten incluir 3 componentes relevantes para el análisis.

La dimensión Fecha permite realizar un análisis en el tiempo de las competencias laborales, pudiendo evidenciar la evolución de la demanda de éstas, siendo un insumo importante al momento de decidir qué competencias laborales requieren esfuerzos para ser incluidas en los perfiles de los postulantes y cuales están quedando obsoletas.

Por otro lado, la dimensión de Puntaje permite segmentar por los grupos o rangos de puntaje, esta característica es necesaria para limpiar el análisis de detecciones de competencias de baja calidad y dejar principalmente el grupo de mayor puntaje que tiene un porcentaje de 95% de detecciones correctas.

El tercer componente de segmentación es aportado por la dimensión Región, que permite realizar el análisis por las regiones del país, lo cual viene a ser relevante para evaluar los esfuerzos regionales invertidos en programas de capacitaciones, ayudando a determinar si estos están en correcta concordancia con las competencias laborales demandadas de dichas regiones.

Finalmente, las dimensiones Competencias y Avisos brindan la oportunidad de segmentar el análisis por las características propias de los distintos mercado o tipos de actividad en cuales se ven involucradas las empresas que publican dichos avisos laborales y características requeridas a los postulantes, como su grado escolar o el tipo de trabajo que ejecutaran.

Capítulo 7

Trabajo Futuro.

Dado el trabajo realizado a lo largo del periodo de desarrollo de esta memoria se ha podido apreciar distintos aspectos que pueden mejorar sus resultados y objetivos de esta memoria, en cuanto a mejoras de desempeño y alcance. Por otro lado, se han encontrado nuevos desafíos que se abren camino a partir de esta primera iteración de un sistema de detección y análisis de demanda de competencias laborales.

A raíz de lo anterior es necesario plasmar estos hallazgos en un capitulo que permita a futuras iteraciones de este trabajo abordar todas estos nuevos desafíos y fronteras. Estos se dividen en 3 puntos importantes; qué se volvería a realizar teniendo nuevas consideraciones, o bien, qué se mejoraría incluyendo nuevos factores, por otro lado, qué nuevos repositorios o qué sistemas se pueden integrar en el proyecto realizado para ampliar el alcance actual de este proyecto y finalmente se proponen nuevos proyectos o ideas que surgen de esta iteración.

7.1. Distintas mejoras al sistema actual.

Luego de haber realizado el proyecto completo se ha concluido y encontrado distintos aspectos que pueden mejorar el desempeño del sistema actual de detección de competencias laborales.

Cantidad de coincidencias encontradas por el sistema de detección.

La cantidad actual de coincidencias encontradas por el sistema rondan los 400 millones, considerando que el repositorio de avisos laborales es de cerca de 680.000 avisos, da como promedio que cada aviso laboral presenta o requiere de aproximadamente 588 competencias laborales distintas en promedio, de un repositorio de solamente 2.150 competencias laborales.

Esto indica que claramente hay una sobre estimación de competencias laborales demandas en un aviso laboral. Para mejorar esta situación y eliminar coincidencias innecesarias se propone utilizar el filtro del motor SOLR para no considerar coincidencias con puntaje entre 0 y 1. Estas son la gran mayoría de las coincidencias y son almacenadas solo por coincidir en alguna pequeña palabra de la competencia y el aviso.

Mejorar la evaluación de la detección de las competencias laborales

En esta iteración se realizó una evaluación de las competencias laborales detectadas, calificando coincidencias a lo largo de la distribución de estos y separando en grupos desde el puntaje cercano a 0 hasta el grupo que consideraba desde el 16 y más alto. Esto permitió discernir y evaluar qué grupo o rangos de puntajes deben ser considerados y cuáles no a lo largo de la distribución de puntajes.

Pero hay un aspecto de los puntajes que puede ser incluido para mejorar esta evaluación. Al momento de realizar una consulta en el servidor SOLR, esta toma en consideración el contexto de la misma, tomando factores como la cantidad de palabras que se buscarán en el repositorio, esta cantidad viene dada por la competencia laboral en sí, en cómo fue descrita en el repositorio original de ChileValora y cuántas palabras necesita para dejar clara su definición.

Este aspecto provoca, por ejemplo, que competencias que posean una descripción corta, de cerca de 4 ó 5 palabras, tenga una distribución de puntajes concentrada en valores más bajos, y a pesar de tener buenas coincidencias, no sean consideradas en los análisis debido a esta distribución.

Entonces se propone realizar una evaluación que no separe por rango de puntajes, sino que ocupe un método que separe los grupos de puntajes por su posición dentro de las coincidencias de la misma competencia.

De esta forma el análisis del rendimiento no sería en rangos absolutos de puntajes, sino que en rangos que dependen de en qué posición respecto a todas las coincidencias de la misma competencia se encuentra.

Migrar a una arquitectura de procesamiento compartido.

A lo largo del desarrollo de este proyecto se presentaron varias situaciones donde había que gastar horas en procesamiento, tanto para utilizar el motor SOLR para la detección de competencias laborales, como para ejecutar el proceso ETL y poblar el modelo estrella que utilizaría finalmente la plataforma de BI.

Estos procesos retrasaban el desarrollo de la memoria, y en la actualidad existen tecnologías para hacer un procesamiento distribuido en distintos computadores, tanto para el motor de búsqueda SOLR, como para el análisis en el data warehouse.

Se propone construir las nuevas de iteraciones sobre estas tecnologías de carga distribuida para tener una mayor capacidad de procesamiento y un menor tiempo de respuesta.

Considerar nuevos campos en el análisis y detección de competencias laborales.

Para la detección de competencias laborales se utilizaron los atributos de los avisos de trabajo Descripción y Otros requisitos, descritos en la Tabla 2.2, estos se juntaban y se utilizaban para el análisis de text mining y la detección de competencias laborales.

Acá surgen dos mejoras, en primer lugar, a veces la gente encargada de publicar el aviso de trabajo duplicaba textualmente la descripción del aviso en estos dos campos lo que causa una sobre estimación al calcular el puntaje asociado a la coincidencia. Por esta razón se propone que al momento de confeccionar el repositorio de avisos laborales se implemente una detección de estas situaciones y solo sean contadas una vez.

En segundo lugar, existe un campo también de texto libre, este es el título del aviso laboral donde el usuario resume su aviso en el nombre del cargo a desempeñar, por ejemplo, "mecánico" o "analista de ventas". Entonces aquí surge una oportunidad para analizar este campo y lograr extraer un perfil laboral o un cargo estandarizado que está buscando el empleador, si esto se consigue, se puede utilizar los perfiles laborales elaborados por ChileValora o utilizar la información ya generada en el sistema para tener una relación o probabilidad entre dichos cargos estandarizados y las competencias laborales, y finalmente es un recurso o un apoyo para encontrar de forma más eficiente las competencias laborales requeridas por los avisos de trabajo.

7.2. Inclusión de nuevos repositorios y métodos.

Para el desarrollo de esta memoria se confeccionaron dos repositorios por separado, el primero incluye los avisos laborales de la BNE entre los años 2011 y mediados de 2016, el segundo repositorio corresponde al catálogo conformado por la comisión ChileValora de competencias laborales, este último incluye 2.150 competencias laborales diferentes.

Ambos repositorios son recursos para el sistema y podrían ser remplazados eventualmente por otros nuevos o de otras fuentes, siempre que se mantenga o se transformen para coincidir en formato con el sistema.

Repositorio de competencias laborales.

El actual repositorio del sistema de competencias laborales corresponde al elaborado por ChileValora, cuenta con 2.150 competencias laborales distintas jerarquizadas por perfil laboral y sector industrial.

Este repositorio presenta ventajas en cuanto a que su contexto de elaboración fue el mercado laboral chileno y que sus elementos son los utilizados para la planificación y aprobación de distintos cursos de capacitaciones por parte del Estado.

Sin embargo, existen otras opciones de repositorios de competencias laborales. En particular, como fue explicado al principio de este informe, la ESCO, Clasificación Europea de competencias, cualificaciones y ocupaciones, elaboró un repositorio de competencias laborales que al momento del desarrollo de esta memoria no estaba disponible.

Esta cuenta con cerca de 10.000 competencias laborales, cada una está asociada a distintas ocupaciones y especificando si son esenciales para la ejecución de dichas ocupaciones o si son optativas. A diferencia del repositorio actual que no supone o no incluye la reutilización o posibilidad de compartir competencias laborales entre distintos cargos u ocupaciones. [39]

Se propone incluir este repositorio en futuras iteraciones, ya que posee más de 4 veces la cantidad actual de competencias laborales y, no menos importante, su posible aplicación post detección de competencias laborales.

Una de esas posibles aplicaciones es elaborar sistemas de recomendaciones laborales para personas con ciertos sets de competencias laborales o para empresas que busquen llenar un cargo y no consideren personas provenientes o con experiencias en cargos que compartan dichas competencias.

Se destaca además que es necesario filtrar aún mejor las coincidencias entre competencias laborales y avisos de trabajo, actualmente hay cerca de 400 millones de coincidencias sobre 2.150 competencias, estimando linealmente, con 10.000 competencias laborales esta cantidad aumentaría a 1.600 millones de coincidencias y ya quedó expuesto en el análisis del rendimiento de los puntajes asociados que los puntajes más bajos no presentan una capacidad efectiva de detección de competencias laborales.

Nuevos repositorios de avisos laborales.

Como se discutió al principio de esta sección, el repositorio de avisos laborales es un recurso del sistema y puede ser remplazado, en este caso la aplicación del sistema es propia del contexto del repositorio, es decir, pertenecen a la BNE y puede ser una herramienta de gestión para los actores involucrados.

Pero en esencia un aviso de trabajo perteneciente a la BNE o a un portal laboral privado no difieren mucho entre sí, ambos poseen un mismo objetivo, un empleador publica el aviso para buscar postulantes y para eso realiza una descripción del trabajo ofrecido, sus requisitos y otros. Fundamentalmente la estructura de los avisos de trabajos en general es la misma independiente del contexto.

Dado lo anterior, existe la opción de integrar al repositorio la información de otros sistemas, como los descritos en la Tabla 2.2, de esta forma se puede realizar un análisis de demanda de competencias laborales más amplio en cuanto al mercado laboral chileno y por otro lado permite acercar el proyecto a nuevos desafíos con índole más comercial que estatal.

7.3. Nuevos desafíos y líneas de investigación.

En esta sección se discute, a partir del trabajo desarrollado, qué nuevas facetas se pueden emplear en las siguientes iteraciones de esta memoria, en cuanto a nuevas aplicaciones del sistema. Por otro lado, también se presentan nuevos trabajos a desarrollar que amplían el alcance o posibles integraciones de sistemas más extensos que cubran otros aspectos del mercado laboral.

Implementar el sistema a tiempo real.

Actualmente el sistema de detección y análisis de competencias laborales se aplica sobre información y repositorios de avisos laborales que datan del año 2011 hasta mediados del 2016, en otras palabras, de avisos laborales pasados y no se implementó ningún tipo de actualización constante de los avisos laborales.

Por esta razón se hace natural proponer una implementación que considere integrar este sistema al portal laboral de la BNE y actualice constantemente los nuevos avisos de trabajo que se vayan publicando en el portal laboral de forma que el análisis sea a tiempo presente.

Esto permitiría al sistema detectar nuevas tendencias en el mercado laboral y servir como una herramienta de gestión que ayuda a realizar correcciones a los programas que se estén implementando en cuanto a capacitación laboral.

Para realizar esto se presentarán varios desafíos tecnológicos, y serán necesarias distintas adaptaciones del sistema actual. En particular, la forma de calcular los puntajes de las coincidencias toma en consideración el contexto del repositorio donde se buscan las competencias laborales, por eso incluye factores como el largo promedio de los elementos del repositorio, factor que cambiaría constantemente al agregar nuevos avisos al repositorio. Entonces es necesario demostrar que, dada una cantidad de avisos de trabajo, este factor converge a un valor determinado.

Hay que tomar en consideración estos factores para determinar si es necesario recalcular los puntajes continuamente, cada cierto tiempo o si finalmente convergerán y no será necesario una actualización de ellos.

Oferta de competencias laborales.

La presente memoria propone y desarrolla un sistema para la detección y análisis de la demanda de competencias laborales, este es un primer paso que naturalmente da paso a medir las competencias de los otros actores, en este caso, a los postulantes o trabajadores que ofrecen una colección de competencias laborales para ejecutar distintos trabajos.

Hay que recordar, como fue explicado en la introducción de esta memoria, que una diferencia entre las competencias laborales que necesitan las empresas y que se encuentran en la fuerza laboral implica distintos costos al mercado en general y al país.

El esfuerzo invertido en poder medir las competencias laborales ofrecidas permitirá, en primer lugar, contrastar esa información con la generada en este trabajo y de esta forma medir la brecha entre ambas y generar acciones para disminuirla.

En segundo lugar, puede surgir una investigación que plantee medir o estimar si estas diferencias afectan a la ejecución de los trabajos laborales, también si son originadas por desconocimiento de las empresas en cuanto a que competencias son necesarias para cada cargo o si los postulantes tienen conocimiento de sus propias competencias.

Evaluar los distintos programas de capacitaciones.

Este sistema pretende ser una herramienta de gestión y apoyo de decisión, a distintos actores gubernamentales ligados a la conformación de programas de capacitaciones y habilitación de empresas para ejercer cursos de capacitaciones.

Entonces un proyecto con mayor impacto sería la implementación de este sistema con esos actores, que se les confeccione distintas métricas y reportes para evaluar si los programas que

estén en curso o los que estén en etapa de planificación responden a las necesidades del mercado en cuanto a competencias laborales.

Un proyecto así debe considerar no solo la tarea tecnológica de elaborar dichos reportes o tableros y brindarles acceso a los involucrados, sino que debe comprender todos los desafíos de una implementación de una herramienta nueva en una cultura organizacional estatal.

Recomendaciones en publicaciones de avisos laborales.

La información generada por el sistema desarrollado permite saber las competencias laborales que necesita cada aviso de trabajo publicado, lo que abre una posible aplicación para los nuevos avisos entrantes al sistema de la BNE.

Se propone desarrollar un sistema que ayude a los usuarios a publicar avisos laborales a considerar competencias laborales requeridas en sus anuncios que no estuviesen agregando a la publicación, todo esto en base a la información que se tenga de avisos laborales de similares características, como área industrial, tipo de cargo, actividad de la empresa, entre otros.

Finalmente se puede completar el proyecto midiendo las postulaciones al cargo y el impacto que tuvo las recomendaciones entregadas, o desde otro punto de vista, haciendo un análisis respecto a la aceptación de los usuarios sobre las competencias laborales recomendadas, investigar si estas no eran consideradas necesarias por el empleador o el sistema de recomendaciones no era adecuado.

Recomendaciones a postulantes en plataformas laborales.

El sistema ha generado información respecto a cada trabajo publicado y las competencias que requieren los empleadores para su ejecución. Es razonable encontrar distintos trabajos o cargos ofrecidos que comparten las competencias laborales requeridas, e incluso se podrían encontrar en rubros e industrias muy diferentes entre sí, como lo pueden ser vendedores en distintos mercados.

Surge la oportunidad de implementar un sistema que, dada las búsquedas de avisos de trabajo de un usuario en una plataforma laboral, se le presenten recomendaciones de cargos que y avisos que compartan competencias laborales sus requeridas.

De esta forma el usuario podría considerar nuevos trabajos en diferentes rubros que antes no consideraba, ayudando al tiempo necesario para encontrar un nuevo trabajo.

Capítulo 8

Conclusiones.

En el presente y último capítulo de este informe de memoria se da a conocer si los objetivos e hipótesis investigativas de esta memoria fueron cumplidos y probados luego del trabajo desarrollado y expuesto en los capítulos anteriores.

Tomando en cuenta que el trabajo realizado tiene como objetivo general, detectar competencias laborales desde avisos de trabajos de la BNE y el desarrollo de una plataforma que habilite un estudio diferenciado por las características de los avisos de trabajos, el contexto de los empleadores y características propias de las competencias laborales. Se debe evaluar bajo estas tareas si fueron cumplidos con el desarrollo de esta memoria.

El primer objetivo de detectar competencias laborales en avisos de trabajos fue logrado a través de la implementación de técnicas de text mining con un motor de búsqueda y la integración de repositorios de competencias laborales elaboradas por ChileValora y avisos de trabajo de la BNE entre los años 2011 y mediados del 2016.

Si bien, de los 400 millones de coincidencias generadas entre las competencias laborales y los avisos de trabajo, su gran mayoría no posee un buen poder de detección, existe un grupo de puntajes más altos que tienen una precisión del 95 % al momento de detectar competencias. Además, la cantidad de coincidencias pertenecientes a este grupo permiten reconocer en promedio 2 competencias laborales por cada aviso de trabajo, no cayendo en una posible sobre estimación de competencias laborales requeridas. Cumpliendo finalmente este primer objetivo general.

El segundo objetivo hace referencia a permitir un análisis segmentado por las características de los avisos laborales, del contexto de las empresas que publican dichos avisos y finalmente por las características de la competencia laboral.

Para alcanzar este objetivo se implementó una plataforma de análisis mediante un data warehouse. En esta plataforma se incluyeron distintos procesos para transformar la información e insertarlas en un modelo estrella que relacionaba todos los componentes para segmentar el análisis con las coincidencias entre competencias laborales y avisos de trabajo.

De esta forma se implementó una plataforma que permite un análisis por distintas características o dimensiones simultáneamente, habilitando a los usuarios para usar esta plataforma desde cualquier navegador web, necesitando solo conexión a internet.

Concluyendo, se asevera que con el uso de esta plataforma se habilita un sistema para analizar las competencias laborales demandadas en los avisos de la BNE segmentando el análisis por características temporales, de sector industrial, de tipo de cargo entre otros. Dando el segundo objetivo de esta memoria como satisfecho.

Además del cumplimiento de los objetivos generales hay otros aspectos que destacar, por un lado, con el desarrollo del trabajo se encontraron diversos aspectos que pueden mejorar el sistema actual, como lo son mejores formas de evaluar el desempeño de la detección de competencias laborales, así como nuevos repositorios para integrar al sistema, que implican un análisis más amplio del mercado laboral.

Finalmente, se recomienda continuar con futuras investigaciones en base a esta memoria, como lo son; una implementación a tiempo real, desarrollar sistemas de recomendaciones a usuarios de portales laborales (tanto empleadores como postulantes) y de esta forma aumentar el impacto de estas investigaciones incluyendo el aspecto de investigación y un posible impacto económico.

Bibliografía.

[1] CHILE. Ministerio de Trabajo y Previsión Social. 2008. Ley 20.267: Crea el sistema nacional de certificación de competencias laborales y perfecciona el estatuto de capacitación y empleo, 25 de junio, 2008.

[2] ESCO: Clasificación europea de capacidades/competencias, cualificaciones y ocupaciones. Skills Pilar: Knowledge, skill and competences, ESCOpedia [en línea] <https://ec.europa.eu/esco/portal/escopedia/Skills_pillar >[consultado: 25 de agosto, 2017]

[3] CAPELLI, P. 2014. Skill Gaps, Skill Shortages and Skill Mismatches: Evidence for the US. Paper. National Bereau of Economic Research.

[4] BRAMER, M. 2016. Introduction to Data Mining. En: Principles of Data Mining 3ª ed. Reino Unido, Portsmouth. Springer. pp. 1-8.

[5] INDURKHYA, N. y DAMERAU, F. J. 2010. Handbook of Natural Language Processing. 2ª ed. Reino Unido, Cambridge. CRC Press.

[6] MANNING, D. RAGHAVAN P. y SCHÜTE, H. 2009. Introduction to Information Retrieval. Inglaterra, Cambridge. Cambridge University Press. 32p.

[7] GENTZKOW M., KELLY B., T. y TEDDY M. 2017. Text as Data. Working Paper. National Bereau of Economic Research.

[8] HEDDEN H. Chapter 1: What are taxonomies? En: The accidental Taxonomist. 2ª ed. Estados Unidos. Information Today, Inc.

[9] Clasificación europea de capacidades/competencias, cualificaciones y ocupaciones. European Skills, Competences, Qualifications and Occupations , ESCOpedia [en línea] <https://ec.europa.eu/esco/portal/escopedia/European_Skills%252C_Competences%252C_Qualifications_and_Occupations_%2528ESCO%2529 > [consultado: 25 de agosto, 2017]

[10] ChileValora, Por qué lo hacemos. [en línea] <http://www.chilevalora.cl/que-hacemos/por-que-lo-hacemos/> [consultado: 25 de agosto, 2017]

[11] NOMINET TRUST. 2012. Employment and the internet. [en línea] <https://www.nominettrust.org.uk/sites/default/files/NT%20SoA%204%20-%20Employment%20and%20the%20internet.pdf > [consultado: 24 de agosto, 2017]

[12] NOMINET TRUST. Our History. [en línea] <https://www.nominettrust.org.uk/who-we-are/our-history> [consultado: 25 de agosto, 2017]

[13] CAPITAL HUMANO. ¿Cuáles son los principales sitios web para buscar empleo en Chile? [en línea] <http://capitalhumano.emol.com/2292/websites-buscar-empleo-en-chile/> [consultado: 25 de agosto, 2017]

[14] SENCE. Bolsa Nacional de Empleo, Antecedentes [en línea] <http://www.sence.cl/601/w3-propertyvalue-510.html?_noredirect=1> [consultado: 25 de agosto, 2017]

[15] INMON, W., H. 2002. Chapter 2: The data warehouse environment. En: Building the Data Warehouse. Estados Unidos. Wiley Computer Publishing.

[16] KIMBALL, R. y ROSS, M. 2002. Chapter 1: Dimensional Modelling Primer. En: The Data Warehouse Toolkit. Estados Unidos. Wiley Computer Publishing.

[17] INTERNATIONAL LABOUR ORGANIZATION. World of Works 2014 Developing with Jobs. Ginebra. PRODOC.

[18] KAMPELMANN S. y RYCX F. 2012. The Impact of Educational Mismatch on Firm Productivity: Direct Evidence from Linked Panel Data. Economics of Education Review.

[19] DELLOITE y THE MANUFACTURING INSTITUTE. 2015. The skills gap in U.S. manufacturing 2015 and beyond [en línea]. <http://www.themanufacturinginstitute.org/~/media/827DBC76533942679A15EF7067A704CD.ashx > [consulta: 20 de abril 2017].

[20] PRAET P. 2015. Lifting potential growth in the euro área [en línea]. European Central Bank. 23 de abril, 2015 <https://www.ecb.europa.eu/press/key/date/2015/html/sp150423.en.html >. [consulta: 20 de abril 2017].

[21] GUVENEN F, KURUSCU B, TANAKA S y WICZER D. Multidimensional Skill Mismatch. Federal Reserve Bank of ST. Louis.

[22] SENCE. Acerca del SENCE. [en línea] <http://www.sence.cl/portal/Acerca-del-Sence/> [consultado: 26 de agosto, 2017]

[23] CHILE. Dirección de Presupuestos 2016. Ley 20.981: Ley de Presupuestos del Sector Público, 15 de diciembre, 2016.

[24] SENCE. Anuarios 2004 a 2016. [en línea] <http://www.sence.cl/portal/Estudios/Anuario-estadistico/Anuarios-2004-a-2016/> [consultado: 26 de agosto, 2017]

[25] CHILEVALORA. Qué hacemos. [en línea] <http://www.chilevalora.cl/que-hacemos/> [consultado: 26 de agosto, 2017]

[26] MINISTERIO DE TRABAJO Y PREVISIÓN SOCIAL. Misión y Objetivos. [en línea] <http://www.mintrab.gob.cl/nuestro-ministerio/mision-y-objetivos/> [consultado: 26 de agosto, 2017]

[27] PORTER, M. F. 1980. An algorithm for suffix stripping. Computer Laboratory, Cambridge, UK.

[28] SNOWBALL. Spanish stemming algorithm. [en línea] <http://snowballstem.org/algorithms/spanish/stemmer.html> [consultado: 30 de septiembre, 2017]

[29] SOLR. Community. [en línea] <http://lucene.apache.org/solr/community.html> [consultado: 2 de noviembre, 2017]

[30] APACHE. Apache Software Foundation. [en línea] <http://www.apache.org/foundation/ > [consultado: 2 de noviembre, 2017]

[31] Chile Valora. Catálogo de competencias laborales. <http://www.chilevalora.cl/buscador/index.php/PerfilCompetencia/index> [consultado: 5 de noviembre, 2017]

[32] SOLR. Learning to Rank. [en línea] <https://lucene.apache.org/solr/guide/6_6/learning-to-rank.html> [consultado: 7 de noviembre, 2017]

[33] SOLR. Overview of searching in SOLR. [en línea] <https://lucene.apache.org/solr/guide/6_6/overview-of-searching-in-solr.html#overview-of-searching-in-solr> [consultado: 8 de noviembre, 2017]

[34] SOLR. Using SorlJ. [en línea] <https://lucene.apache.org/solr/guide/6_6/using-solrj.html> [consultado: 9 de noviembre, 2017]

[35] SNOWBALL. Introductiong. [en línea] < http://snowballstem.org > [consultado:11 de noviembre, 2017]

[36] SAIKU. Introduction. [en línea] <http://saiku-documentation.readthedocs.io/en/latest/> [consultado: 12 de noviembre, 2017]

[37] PENTAHO. What is a schema? [en línea] <https://mondrian.pentaho.com/documentation/schema.php#What_is_a_schema> [consultado: 15 de noviembre, 2017]

[38] MICROSOFT. What Are The Data Warehouse OLAP Cubes? [en línea] <https://msdn.microsoft.com/en-us/library/bb219339(v=cs.70).aspx> [consultado: 17 de noviembre, 2017]

[39] ESCO. Herramientas y Recursos. [en línea] <https://ec.europa.eu/esco/portal/version> [consultado: 29 de noviembre]

[40] CHILEVALORA. Informe de ejecución presupuestaria. [en línea] <http://www.chilevalora.cl/transparencia/2016/presupuesto.html > [consultado: 4 de diciembre]

[41] SOLR REF GUIDE 6.6. Shingle Filter. [en línea] <https://lucene.apache.org/solr/guide/6_6/filter-descriptions.html#FilterDescriptions-ShingleFilter> [consultado: 6 de enero]

[42] INMON, W., H. 2002. Chapter 3: The data warehouse and design. En: Building the Data Warehouse. Estados Unidos. Wiley Computer Publishing.

[43] MARIADB. Why MariaDB?. [en linea] <https://mariadb.com/products/technology/server> [consultado: 6 de enero]

[44] SENCE. Franquicia Tributaria de Capacitación [en linea] <http://www.sence.cl/601/w3-printer-583.html > [consultado: 7 de enero]

[45] CHILE. Ministerio de Trabajo y Previsión Social. 2001. Ley 19.728: Establece un seguro de desempleo, 14 de mayo, 2001.

[46] MICROSOFT. Manipulate and Query OLAP Data using ADOMD and Multidimensional Expression. [en línea] <https://www.microsoft.com/msj/0899/mdx/mdx.aspx> [consultado: 7 de enero]

[47] MICROSOFT. MDX Query - The Basic Query [en línea] <https://docs.microsoft.com/en-us/sql/analysis-services/multidimensional-models/mdx/mdx-query-the-basic-query> [consultado: 7 de enero]

[48] BANFI S., CHOI S., GALEGUILLOS C., VILLENA B. 2017. Informe sobre la Bolsa Nacional de Empleo 2011-2016. Depto. Ingeniería Industrial, Universidad de Chile, Chile.

Anexos.

A. Stopwords en español.

de	son	habia	nosotros	vuestras	estaríais	ha	habíamos	seas	fuisteis	tendreis
la	entre	ante	mi	esos	estariais	hemos	habiamos	seamos	fueron	tendrán
que	está	ellos	mis	esas	estarían	habéis	habíais	seáis	fuera	tendran
el	esta	e	tú	estoy	estarian	habeis	habiais	seais	fueras	tendría
en	cuando	esto	te	estás	estaba	han	habían	sean	fuéramos	tendrías
y	muy	mí	ti	está	estabas	haya	habian	seré	fueramos	tendríamos
a	sin	mi	tu	esta	estábamos	hayas	hube	sere	fuerais	tendríais
los	sobre	antes	tus	estamos	estabamos	hayamos	hubiste	serás	fueran	tendrían
del	ser	algunos	ellas	estáis	estabais	hayáis	hubo	seras	fuese	tenía
se	tiene	qué	nosotras	estais	estaban	hayais	hubimos	será	fueses	tenías
las	también	que	vosotros	están	estuve	hayan	hubisteis	sera	fuésemos	teníamos
por	tambien	unos	vosotras	esté	estuviste	habré	hubieron	seremos	fueseis	teníais
un	me	yo	os	este	estuvo	habre	hubiera	seréis	fuesen	tenían
para	hasta	otro	mio	estés	estuvimos	habrás	hubieras	sereis	siendo	tuve
con	hay	otras	MIO	estemos	estuvisteis	habras	hubiéramos	serán	sido	tuviste
no	donde	otra	mía	estéis	estuvieron	habrá	hubieramos	seran	tengo	tuvo
una	han	él	mia	esteis	estuviera	habra	hubierais	sería	tienes	tuvimos
su	quien	el	míos	estén	estuvieras	habremos	hubieran	seria	tiene	tuvisteis
al	están	tanto	mios	esten	estuviéramos	habréis	hubiese	serías	tenemos	tuvieron
es	estan	esa	mías	estaré	estuvieramos	habreis	hubieses	serias	tenéis	tuviera
lo	estado	estos	mias	estare	estuvierais	habrán	hubiésemos	seríamos	teneis	tuvieras
como	desde	mucho	tuyo	estarás	estuvieran	habran	hubiesemos	seriamos	tienen	tuviéramos
más	todo	quienes	tuya	estaras	estuviese	habría	hubieseis	seríais	tenga	tuvierais
mas	nos	nada	tuyos	estará	estuvieses	habria	hubiesen	seriais	tengas	tuvieran
pero	durante	muchos	tuyas	estara	estuviésemos	habrías	habiendo	serían	tengamos	tuviese
sus	estados	cual	suyo	estaremos	estuviesemos	habrias	habido	serian	tengáis	tuvieses
le	todos	sea	suya	estaréis	estuvieseis	habríamos	habida	era	tengais	tuviésemos
ya	uno	poco	suyos	estareis	estuviesen	habriamos	habidos	eras	tengan	tuvieseis
o	les	ella	suyas	estarán	estando	habríais	habidas	éramos	tendré	tuviesen

fue	ni	estar	nuestro	estaran	estado	habriais	soy	eramos	tendre	teniendo
este	contra	haber	nuestra	estaría	estada	habrían	eres	erais	tendrás	tenido
ha	otros	estas	nuestros	estaria	estados	habrian	es	eran	tendras	tenida
sí	fueron	estaba	nuestras	estarías	estadas	había	somos	fui	tendrá	tenidos
si	ese	estamos	vuestro	estarias	estad	habia	sois	fuiste	tendra	tenidas
porque	eso	algunas	vuestra	estaríamos	he	habías	son	fue	tendremos	tened
esta	había	algo	vuestros	estariamos	has	habias	sea	fuimos	tendréis	

Tabla 8.1: Listado de Stopwords.
Fuente: Snowballstem.

B. Esquema Mondrian v.4

```
    <?xml version="1.0"?>
    <Schema name="Competencias Laborales v2"  metamodelVersion="4.0">
        <PhysicalSchema>
            <Table name="competencia" />
            <Table name="aviso" />
            <Table name="fact" />
            <Table name="grupo" />
            <Table name="region" />
            <Table name="fecha" />
        </PhysicalSchema>
        <Cube name="Competencias Laborales v2">
            <Dimensions>
                <Dimension name="Avisos" table="aviso" key="AvisoID">
                    <Attributes>
                        <Attribute name="Area" column='area' hierarchyHasAll="true"  >
                            <Key><Column name="area" /></Key>
                        </Attribute>
                        <Attribute name="Tipo de Cargo" column='cargo_general' hierarchyHasAll="true" >
                            <Key><Column name="cargo_general" /></Key>
                        </Attribute>
                        <Attribute name="Grado Escolar" column='grado_escolar' hierarchyHasAll="true" >
                            <Key><Column name="grado_escolar" /></Key>
                        </Attribute>
                        <Attribute name="Tipo de Actividad" column='actividad' hierarchyHasAll="true" >
                            <Key><Column name="actividad" /></Key>
                        </Attribute>
                        <Attribute name="Disponibilidad de Horario" column='disponibilidad'
hierarchyHasAll="true" >
                            <Key><Column name="disponibilidad" /></Key>
                        </Attribute>
                        <Attribute name="Situacion de Estudio" column='situacion_estudio'
hierarchyHasAll="true" >
                            <Key><Column name="situacion_estudio" /></Key>
                        </Attribute>
                        <Attribute name="AvisoID" column='idaviso' >
                            <Key><Column name="idaviso" /></Key>
                        </Attribute>
                    </Attributes>
                </Dimension>
                <Dimension name="Competencia" table="competencia" key="CompetenciaId">
                    <Attributes>
                        <Attribute name="Sector" column='sector' hierarchyHasAll="true"  >
                            <Key><Column name="sector" /></Key>
                        </Attribute>
                        <Attribute name="Contenido" column="competencia" >
                            <Key><Column name="competencia" /></Key>
                        </Attribute>
                        <Attribute name="CompetenciaId" column='idcompetencia' >
                            <Key><Column name="idcompetencia" /></Key>
                        </Attribute>
                    </Attributes>
                </Dimension>
                <Dimension name="Puntaje" table="grupo" key="grupoID">
                    <Attributes>
                        <Attribute name="Rango" column="rango" orderByColumn="idgrupo"  >
                            <Key><Column name="rango"/></Key>
                        </Attribute>
                        <Attribute name="grupoID" column="idgrupo" orderByColumn="idgrupo"  >
                            <Key><Column name="idgrupo"/></Key>
                        </Attribute>
```

Ilustración 8.1: Esquema Mondrian PARTE 1.
Fuente: Elaboración propia.

```
    <Dimension name="Region" table="region" key="regionID">
                    <Attributes>
                        <Attribute name="Region" column="nombre" orderByColumn="nombre"  >
                            <Key><Column name="nombre"/></Key>
                        </Attribute>
                        <Attribute name="regionID" column="idregion" orderByColumn="idregion" >
                            <Key><Column name="idregion"/></Key>
                        </Attribute>
                    </Attributes>
                </Dimension>
                <Dimension name="Fecha" table="fecha" key="fechaID">
                    <Attributes>
                        <Attribute name="fechaID" column="idfecha" orderByColumn="idfecha"  >
                            <Key><Column name="idfecha"/></Key>
                        </Attribute>
                        <Attribute name="Año" column="year" orderByColumn="year"  >
                            <Key><Column name="year"/></Key>
                        </Attribute>
                        <Attribute name="Mes" column="month" orderByColumn="month" >
                            <Key><Column name="month"/></Key>
                        </Attribute>
                        <Attribute name="Dia" column="day" orderByColumn="day" >
                            <Key><Column name="day"/></Key>
                        </Attribute>
                        <Attribute name="Quarter" column="quarter" orderByColumn="quarter" >
                            <Key><Column name="quarter"/></Key>
                        </Attribute>
                        <Attribute name="Semana" column="week" orderByColumn="week" >
                            <Key><Column name="week"/></Key>
                        </Attribute>
                    </Attributes>
                </Dimension>
            </Dimensions>
            <MeasureGroups>
                <MeasureGroup name="Competencias Laborales y Avisos" table ="fact">
                    <Measures>
                        <Measure name="Cantidad de Competencias Laborales" column="idfact" aggregator =
"count" />
                        <Measure name="Cantidad Avisos" column="aviso_idaviso" aggregator = "distinct-
count" />
                    </Measures>
                    <DimensionLinks>
                        <ForeignKeyLink dimension="Competencia"
foreignKeyColumn="competencia_idcompetencia"/>
                        <ForeignKeyLink dimension="Puntaje" foreignKeyColumn="grupo_idgrupo"/>
                        <ForeignKeyLink dimension="Region" foreignKeyColumn="region_idregion"/>
                        <ForeignKeyLink dimension="Avisos" foreignKeyColumn="aviso_idaviso"/>
                        <ForeignKeyLink dimension="Fecha" foreignKeyColumn="fecha_idfecha"/>
                    </DimensionLinks>
```

Ilustración 8.2: Esquema Mondrian PARTE 2.
Fuente: Elaboración propia.