



UNIVERSIDAD
NACIONAL
DE COLOMBIA

Desarrollo de un Sistema de recomendaciones para el fortalecimiento de competencias laborales a profesionales recién egresados usando Ontologías y RDFs

Luis Henry Andrade Mosquera

Universidad Nacional de Colombia
Facultad de Ingeniería, Departamento de Ingeniería de Sistemas e Industrial,
Bogotá, Colombia

2018

Desarrollo de un Sistema de recomendaciones para el fortalecimiento de competencias laborales a profesionales recién egresados usando Ontologías y RDFs

Luis Henry Andrade Mosquera

Trabajo de investigación presentado como requisito parcial para optar al título de:
Magíster en Ingeniería - Ingeniería de Sistemas y Computación

Director (a):

MSc. Ing. Carlos Hernán Caicedo Escobar Caicedo

Línea de Investigación:

Datos Enlazados, RDF, Ontología y web semántica.

Universidad Nacional de Colombia

Facultad de Ingeniería, Departamento de Ingeniería de Sistemas e Industrial,
Bogotá, Colombia

2018

Dedico este trabajo a todas aquellas personas que con su confianza me comprometen a no renunciar, a procurar ser mejor cada día, a convertir mis flaquezas en fortalezas, a todos ellos les debo mi gratitud.

“¿Quién puede compararse al sabio? ¿Quién conoce el sentido de las cosas? La sabiduría ilumina la cara del hombre; hace que cambie su duro semblante.”

Eclesiastés 8:1

Agradecimientos

Quiero agradecer a Dios ante todo por permitirme pertenecer a una de las comunidades académicas más destacadas del país y a mantener viva la esperanza en conseguir este importante logro; a mis padres Luis y Ludivia por inculcar los valores que tanto en lo académico como en lo personal me han permitido afrontar retos y superar dificultades sin renuncias o afectaciones dadas por las circunstancias, a mi hermano Edwin porque su apoyo y contribuciones fueron vitales para llegar hasta este punto en mi carrera profesional, a mi esposa e hijos Sergio y Esteban quienes me llenan diariamente de orgullo y honra.

A mi director de tesis, MSc. Ing. Carlos Hernán Caicedo, por sus valiosos aportes a lo largo de mi trayectoria profesional que me han ayudado en la ejecución de múltiples proyectos en el área de los Sistemas de Información, orientados a la optimización de los procesos de Gestión y seguimiento de Proyectos de extensión e investigación universitaria como también en el efectivo relacionamiento entre la Universidad Nacional de Colombia institución y sus egresados a través de la coordinación del programa de egresados, así mismo al MSc. Ing. Mario Armando Rosero Muñoz que indujo a la forma de estructurar los datos de manera semántica en el presente trabajo.

Por último, quiero agradecer a la universidad pública del país por que canalizan de manera seria la voluntad de una sociedad que, a pesar de todo le sigue apostando a la educación de calidad sin más condición que el mérito de sus estudiantes.

Resumen

El Presente trabajo emplea las funcionalidades relativas a algunos componentes de la Web semántica, para habilitar interfaces Web que permiten mediante búsquedas SPARQL, obtener información de estudios complementarios a los perfiles de egresados recién graduados almacenados en la base de datos del sistema de información de egresados de la Universidad Nacional de Colombia. Dichos estudios complementarios serán recomendados mediante el procesamiento de transacciones almacenadas del módulo de bolsa de empleo, en el que los egresados más antiguos logran ser vinculados a procesos de contratación por parte de empresas que requieren cubrir vacantes a través del mismo sistema, luego de actualizar sus perfiles con trayectoria académica, experiencia laboral y perfil ocupacional. Este último elemento requiere de un etiquetado basado en herramientas de extracción y anotación en lenguaje natural, con el que se obtienen tripletas que permite búsquedas de perfiles afines a temáticas particulares demandadas por las empresas.

Palabras clave: Búsqueda semántica, anotación semántica, procesamiento del lenguaje natural, Ontologías RDF, SPARQL, Web Semántica.

Abstract

The present work uses the features relating to some components of the Semantic Web, to enable Web interfaces that allow searches SPARQL, information of complementary studies to the newly graduated alumni profiles stored in the database of the information system of graduates from the Universidad Nacional de Colombia. These studies will be recommended by the stored transaction processing module of employment, in which the oldest graduates can be linked to recruitment processes on the part of businesses that require filling vacancies through the same system, then update your profiles with academic career, work experience and occupational profile. This last item requires a labeling based on extraction and annotation tools in natural language, with the obtained triplets that allows searches of profiles related to particular themes demanded by companies.

Keywords: Semantic search, semantic annotation, Natural Language Processing, , RDF Ontologies, SPARQL, Semantic Web.

Contenido

	Pág.
Resumen	VIII
Abstract	IX
Lista de abreviaturas	XVI
Introducción	1
1. Marco General	3
1.1. Antecedentes.....	3
1.1.1. Disponibilidad y aprovechamiento de la información.....	5
1.2. Descripción del problema.....	6
1.3. Objetivos.....	9
1.4. Alcances	9
1.5. Limitaciones.....	10
1.6. Metodología.....	10
2. Marco Teórico	12
2.1 Calidad en la educación y empleabilidad.....	12
2.2 Currículo.....	20
2.2.1. Perfil profesional.....	20
2.2.1.1. Competencias laborales generales	21
2.2.1.2. Competencias laborales específicas	21
2.3 Mercado de trabajo profesional.....	21
2.4 Necesidades del mercado laboral.....	22
2.5 Desajustes del mercado laboral entre la oferta y la demanda	23
2.6 Pertinencia en la formación.....	23
2.7 Formación Continua	25
2.8 Web Semántica	26
3 Estado del arte	27
3.1 Estándares	29
3.1.1 XML.....	29
3.1.2 XSD.....	30
3.1.3 RDF / RDFS.....	30
3.1.4 SKOS	30
3.1.5 SKOS básico	31
3.1.6 SKOS avanzado	31
3.1.7 OWL Lenguajes de Ontologías Web.....	31

3.1.8	OWL Lite	31
3.1.9	OWL DL	32
3.1.10	OWL FULL	32
3.1.11	Lenguaje de Reglas de la Web Semántica (SWRL)	33
3.1.12	RIF Formato de intercambio de reglas	33
3.1.13	SPARQL.....	33
3.2	Conceptos.....	34
3.2.1	Datos Enlazados.....	34
3.2.2	Ontologías o Vocabularios.....	35
3.3	Componentes.....	35
3.3.1	Editor de Ontologías Protege	35
3.3.2	Contenedores de RDF	36
3.3.3	Lenguajes de Consulta	36
3.3.4	Sistemas para extracción de información.....	37
3.3.5	End-point SPARQL, Punto final de acceso	38
3.4	Clasificaciones de sistemas de recomendación	39
3.4.1	Basados en contenido	39
3.4.2	Colaborativos	40
3.4.3	Híbridos.....	40
4 Desarrollo de un Sistema de recomendaciones para el fortalecimiento de competencias laborales a profesionales recién egresados usando Ontologías y RDFs		41
4.1	Esquema de propuesto para el flujo de datos.....	41
4.2	Conversión de datos relacionales a tripletas RDB2RDF	42
4.3	Texto a RDF.....	43
4.3.1	Extracción de datos de perfiles seleccionados por empresas oferentes	46
4.3.2	Obtención de RDFs de perfiles y almacenamiento de etiquetas	48
4.4	Modelado de datos RDF.....	50
4.5	Interfaces	51
4.5.1	SPARQL EndPoint.....	51
4.5.2	Formulario de consultas.....	52
4.6	Resultados de las pruebas	53
4.6.1	Evaluación experimental.....	55
5	Conclusiones y recomendaciones	58
5.1	Conclusiones	58
5.2	Recomendaciones	59

Lista de figuras

Figura 1-1: Avances significativos que en las últimas dos décadas han contribuido al estado del arte actual de la tecnología de datos enlazados.....	4
Figura 1-2: Módulo de gestión de hoja de vida del Sistema de Información de Egresados.	7
Figura 1-3: Comunidad Red de Prestadores de Servicio Público de Empleo.....	8
Figura 2-1: Arquitectura del Sistema de Aseguramiento de la Calidad de la Educación Superior.....	17
Figura 2-2: Entradas y salidas del Sistema O3.....	18
Figura 2-3: Elementos básicos del Sistema Nacional de Formación Técnica	24
Figura 2-4: Ejemplo del bloque básico de construcción de una tripleta RDF	30
Figura 4-1: Esquema de propuesto para el flujo de datos.....	42
Figura 4-2: Consulta realizada con utilidad Query designer de Oracle APEX, sobre el modelo ER de la BD transaccional que define la vista usada para el proceso RDB2RDF.	43
Figura 4-3: Estructograma del método de parametrización vía Post createPostMethod ()	44
Figura 4-4: Estructograma del método de selección de archivos run ()	44
Figura 4-5: Estructograma del método saveResponse (File file, PostMethod method) ..	45
Figura 4-6: Modelo UML de la clase cliente que invoca el Servicio Web de etiquetado automático.....	45
Figura 4-7: Flujo para la extracción de datos de perfiles a ficheros en formato .xls	46
Figura 4-8: Flujo para la obtención de un único archivo de texto para cada uno de los perfiles.	47
Figura 4-9: Flujo para la obtención y almacenamiento de las etiquetas obtenidas en el RDF.....	49
Figura 4-10: Representación gráfica de la ontología adaptada.....	50
Figura 4-11: Vista del SPARQL EndPoint del servidor SPARQL Apache Jena Fuseki 2.4 implementado.....	51
Figura 4-12: Salida de una consulta de prueba del SPARQL EndPoint implementado. .	52
Figura 4-13: Diagrama de caso de uso para la interacción con la interfaz de visualización de recomendaciones.	52

Figura 4-14: Interfaz para la selección de programas académicos para obtención de recomendaciones.....	53
Figura 4-15: Interfaz para la selección de temáticas específicas para la obtención de perfiles opcionados.	54
Figura 4-16: Interfaz que le permite al gestor de ofertas autorizado verificar el perfil resultante de la consulta.	54
Figura 4-17: Resultados de caso de prueba para atender solicitud de empresa.	56
Figura 4-18: Resultados de caso de prueba para atender solicitud de empresa – Vista detallada.	56
Figura 4-19: Resultados de caso de prueba para atender solicitud de egresado.	57

Lista de tablas

Tabla 3-1 Comparativo de sistemas y sus fuentes RDF para experiencias similares.	27
Tabla 3-2 Resumen de componentes, estándares y conceptos centrales del estado del arte.....	28
Tabla 3-3 Gestores de almacenamiento o contenedores de RDF.	36
Tabla 4-1 Descripción detallada del flujo de datos para la preparación de los datos de perfiles.	46
Tabla 4-2 Descripción detallada del flujo de datos para el preprocesamiento de los datos de perfiles.....	48
Tabla 4-3 Descripción detallada del flujo de datos para el almacenamiento de etiquetas.	49
Tabla 4-4 Actividades de servicio y otras especificaciones operativas.	55

Lista de abreviaturas

Abreviatura	Término
W3C	World Wide Web Consortium
XML	eXtensible Markup Language (World Wide Web Consortium et al., 2013)
XSD	XML Schema Definition Language
RDF	Resource Description Framework
SKOS	Simple Knowledge Organization System
OWL	Web Ontology Language
RIF	Rule Interchange Format
SPARQL	SPARQL Protocol and RDF Query Language
JDK	Java Development Kit
DBMS	Database management system

Introducción

La importancia que se ha venido dando a los temas laborales en Colombia pasa por el fortalecimiento de la economía, la ampliación de la base productiva, la creación de empleos dignos y la formalización de los existentes; al igual que por consolidación del sistema de inspección del trabajo, mecanismos eficientes de conciliación prejudicial del empleo en Colombia, Políticas para corregir la informalidad laboral, todos estos aspectos alentados por la iniciativa del gobierno nacional de ingresar al grupo de naciones miembro de la Organización para la Cooperación y el Desarrollo Económicos OCDE.

Los lineamientos dados por la rama ejecutiva colombiana se han venido complementando por el legislativo con la expedición de normativas como la ley 1636 de 2013, por medio de la cual se adoptan mecanismos de protección al cesante con la finalidad de articular y ejecutar un sistema integral de políticas de mitigación de los efectos del desempleo que afectan a los trabajadores; al tiempo que facilitar la reincorporación de la población cesante en el mercado laboral en condiciones de dignidad, permanencia, formalización y mejoramiento de la calidad de vida.(Congreso de Colombia, 2013)

La universidad Nacional de Colombia dentro del cumplimiento de su misión; al igual que en la implementación de las políticas gubernamentales de mejoramiento de las competencias laborales, y seguimiento a los egresados dentro del Programa de Acreditación Institucional, realiza esfuerzos que contribuyen a los desafíos de empleabilidad que enfrentan los egresados, vinculándose institucionalmente como bolsa de empleo ante la Unidad del Servicio Público de Empleo del Ministerio de Trabajo, comprometiendo recursos humanos e informáticos a través del programa de egresados de la Institución para la atención de los egresados que son producto de su actividad misional y que afrontan el mundo laboral.

Actualmente la bolsa de empleo de la Universidad Nacional brinda servicios básicos reglamentados por el ministerio de trabajo (Ver anexo 1) que le permite gestionar oferentes, ofertas laborales y demandantes de empleos, propiciando que estos se contacten a través de los distintos procesos empresariales de selección laboral. Con el resultado del presente trabajo se le permitirá adicionalmente a dicha dependencia, obtener información útil para ser proporcionada a los recién egresados como recomendación de formación o capacitación adicional que les permita mejorar sus competencias para ser convocados a procesos de selección laboral en la titulación respectiva.

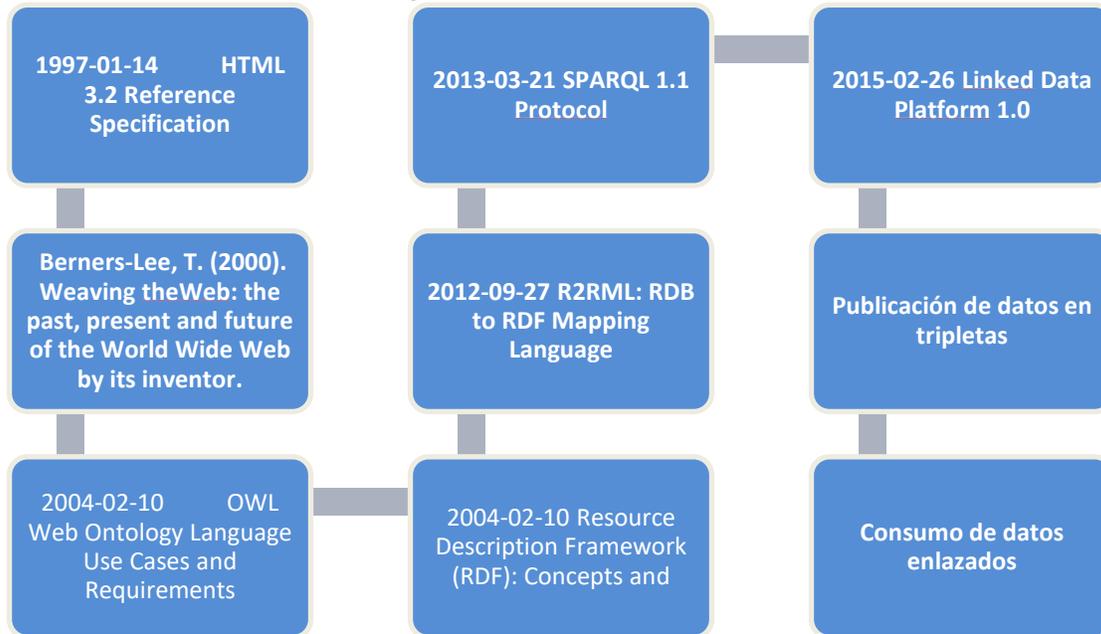
1. Marco General

1.1. Antecedentes

Los esfuerzos de estandarización son tan antiguos como la coordinación de la capacidad de los seres humanos; existen estándares grabados en pictogramas en los monumentos funerarios del antiguo Egipto; en el código de Hammurabi y en el antiguo testamento; y modernamente la Asociación de Metalúrgicos Franceses adelantó acciones para estandarizar tanto los elementos de los que se hacían los trenes como su funcionamiento. Sin olvidar que fue hace un poco más de 25 años cuando Tim Berners Lee quiso reformular la manera en que usamos la información en la edad contemporánea, la forma en la que trabajamos mediante ese invento denominado World Wide Web, que surgió básicamente por la frustración de un ingeniero de software que en un laboratorio enorme y lleno de personas de todo el mundo usaban todo tipo de computadoras, formatos y sistemas de documentación diferentes por lo cual si se quería construir algo en medio de toda esa diversidad, surgía la necesidad de conectarse a nuevos dispositivos, que requerían aprender a ejecutar algún programa nuevo, en algún nuevo formato para poder encontrar la información requerida porque eran todos incompatibles. La frustración por todo este potencial desaprovechado, por todos esos documentos que, si se los imaginaba como parte de algo más grande, de un sistema de documentación virtual en algún sitio, entonces la vida sería mucho más fácil. La idea que inicialmente fue escrita en un memorando sugiriendo el sistema global de hipertexto, y solo 18 meses después se permitió el desarrollo de la idea como una especie de proyecto lúdico, donde se delineó lo que debería ser el HTML, el protocolo de hipertexto HTTP, que nombra las cosas con URLs y que para ese momento no era fácil explicar cómo sería la Web, era difícil hacer que imaginaran como un enlace podría apuntar prácticamente a cualquier documento vinculado que se alentó a poner en la web, es eso lo que no podía visualizarse intuitivamente en ese momento, de manera curiosa el momento actual es similar con la

web de "datos enlazados" que ahora requiere que se pongan datos de casi todo lo que puedan imaginarse en la web de manera masiva y sistemática.

Figura 1-1: Avances significativos que en las últimas dos décadas han contribuido al estado del arte actual de la tecnología de datos enlazados.



Fuente: Elaboración propia

La tecnología de datos enlazados habrá de reevaluar si las direcciones HTTP se están usando no sólo para documentos, para cosas a las que se refieren los documentos, para personas, lugares, productos, eventos, etc. Ya que todo tipo de conceptos, ahora tienen nombres que empiezan con HTTP. Un segundo aspecto a revisar, es si se obtiene información importante, si se toma una de esas HTTP y se busca en la web recupero datos relevantes, en un formato estándar y con información útil que alguien podría querer conocer al respecto. Por ejemplo, sobre cierto suceso saber quién intervino en el mismo y a su vez cualquier cosa sobre esa persona, lugar de nacimiento, entre otros.

Un tercer aspecto es que cuando se obtiene esa información no es sólo caracterizada con ciertos atributos sino con sus relaciones, de esta forma si un registro que detalla que una persona nació en Berlín, la relación inmediata es que Berlín está en Alemania y si este país tiene una relación con una dirección HTTP será posible avanzar en la búsqueda y consultarla.

1.1.1. Disponibilidad y aprovechamiento de la información

Al observar la apertura de datos de gobierno, un hecho concreto en el que el presidente de los Estados Unidos Barack Obama, dijo en un discurso que los datos del gobierno estadounidense deberían estar disponibles en Internet en formatos accesibles. Al nivel de accesibilidad para llegar a los datos de manera enlazada cobra mayor relevancia y trasciende a la premisa de la transparencia porque esos datos son de todos los organismos del gobierno que reflejan cómo se vive en la nación que comparte dicha información, facilitando que el funcionamiento de entidades tanto gubernamentales como las que las rodean. De hecho, si uno conoce algunos datos de un organismo de gobierno, usualmente encuentra que se está muy tentado a no facilitar la información, las entidades se aferran la información que en el mejor de los casos solo se divulga de manera sumariada en algún sitio web. Lo que obliga a formalizar por parte de estados como Colombia la forma en que se dará acceso a dicha información, como es la Ley 1712 de 2014, por medio de la cual se crea la Ley de Transparencia y del Derecho de Acceso a la Información Pública Nacional y se dictan otras disposiciones como son los principios de transparencia, de buena fe, de facilitación, de no discriminación, de gratuidad, de celeridad, de eficacia, de la calidad de la información, de la divulgación proactiva de la información y de la divulgación proactiva de la información.

La mencionada ley se detiene en múltiples definiciones que garanticen la privacidad de los individuos pero además en la de datos abiertos, que son todos aquellos datos primarios o sin procesar, que se encuentran en formatos estándar e interoperables que facilitan su acceso y reutilización, los cuales están bajo la custodia de las entidades públicas o privadas que cumplen con funciones públicas y que son puestos a disposición de cualquier ciudadano, de forma libre y sin restricciones, con el fin de que terceros puedan reutilizarlos y crear servicios derivados de los mismos. Y en consonancia habilitó un catálogo de datos abiertos donde las entidades según su ámbito publican los datos que consideran pueden cumplir con los preceptos dados anteriormente.

El tipo de datos que se espera sean publicados son todos aquellos que no estén sometidos a reserva de acuerdo con la constitución y la ley. Los datos disponibles tendrán, por tanto, una diversidad considerable que abarcará múltiples sectores o temáticas. Ley 1450 de 2011 Art. 230. “Todas las entidades de la administración pública deberán adelantar las acciones señaladas por el Gobierno Nacional a través del Ministerio de las Tecnologías de la Información y las Comunicaciones para la estrategia de Gobierno en Línea. Esta estrategia contemplará como acciones prioritarias estimular el desarrollo de servicios en línea del Gobierno por parte de terceros basados en datos públicos”.

De otro lado durante los últimos años el Gobierno Nacional ha reconocido la importancia del seguimiento a la inserción y desempeño de las personas en el contexto laboral y profesional. Este reconocimiento ha significado la regulación de la manera en que se publican ofertas laborales por parte de las empresas a través de bolsas de empleo cajas de compensación como uno de los ejes que contribuyan a iniciativas de primer empleo.

Las bolsas laborales orientan a los usuarios de las mismas en cuanto a los aspectos procedimentales o en la forma en que se postulan a las disputadas ofertas, sin embargo no tienen facilidades para orientar en cuanto a la formación complementaria adicional para mejorar las posibilidades de obtener un nuevo empleo debido al aislamiento de los datos que se tienen en los sistemas de gestión de ofertas y su escaso nivel análisis que en ocasiones se limita a cifras de frecuencias derivadas de conteos básicos.

1.2. Descripción del problema

La obtención de primer empleo por parte de recién egresados del sistema de educación superior universitaria en Colombia se destaca como el primero de muchos retos que han de ser afrontados por estos en su desempeño profesional, a lo anterior contribuye el gobierno mediante incentivos de tipo legislativo, como la ley 1429 de 2010 estimula la contratación de jóvenes por parte de las empresas a cambio del favorecimiento para las empresas en sus condiciones de operación.

Para el caso concreto de los recién graduados de la Universidad Nacional de Colombia actualmente la Institución acompaña este proceso mediante la vinculación del Programa

de Egresados al servicio público de empleo a través del registro de la institución al directorio de bolsas de empleo reglamentadas por la Unidad Administrativa Especial del Servicio Público de Empleo UASPE.

Hasta hace poco el programa de egresados contaba con una línea de acción muy concreta la consiste en consolidar y enriquecer la base de datos de egresados a lo largo de su trayectoria de ya completa 160 años de antigüedad. La base de datos hasta ahora cuenta con más de 110.000 Egresados a los que se les habilita el acceso al Sistema de Información de Egresados SIE mediante el cual y luego de actualizar su currículum pueden consultar el módulo de ofertas laborales, que corresponde al servicio de bolsa de empleo reglamentado por la Unidad Administrativa Especial del Servicio Público de Empleo UASPE.

Figura 1-2: Módulo de gestión de hoja de vida del Sistema de Información de Egresados.

INFORMACIÓN DEL EGRESADO

DATOS PERSONALES Editar

Datos de Identificación: CÉDULA No. [REDACTED], Masculino

Datos de Contacto: [REDACTED]@unal.edu.co - Personal - Correo de correspondencia

EDUCACIÓN Editar

Estudios Realizados: **INGENIERO EN ... - (PRINCIPAL)**
INGENIERIA INDUSTRIAL - PREGRADO
UNIVERSIDAD NACIONAL DE COLOMBIA
- SEP 15, 2006

Idiomas: **ESPAÑOL, CASTELLANO**
INGLÉS

INFORMACIÓN LABORAL Editar

Experiencia Laboral: **CARGO - INGENIERO INDUSTRIAL, LOGÍSTICA DE PRODUCCIÓN Y DISTRIBUCIÓN (ACTUAL)**
LABORATORIOS OPHALAC S.A
JAN 16, 2006 - (ACTUAL)

Perfil: Soy un profesional con experiencia en Planeación de la Producción, Análisis de Costos de Producción y Control y Administración de Inventarios. Con gran dominio de herramientas de Ofimática, es especial Excel (Funciones avanzadas y programación en visual basic), soy una persona con facilidad para trabajar en equipo y soportar las tareas bajo presión.

← GENERAR VERSIÓN PDF

Get Adobe Reader

Fuente: www.egresados.unal.edu.co

La UASPE que surge como una iniciativa gubernamental vincula cajas de compensación, empresas dedicadas a la intermediación laboral y otras entidades como Universidades que pueden realizar la búsqueda de perfiles adecuados y remitir las hojas de vida que correspondan a los perfiles solicitados por las empresas o sencillamente moderan la interacción entre las empresas y los egresados, garantizando unos mínimos de información y/o calidad de la misma, este último es el esquema de servicio autorizado (ver anexo 1) con que cuenta la bolsa de empleo de la Universidad y que se renueva cada año.

Con el fin de mejorar la orientación que se le pueda brindar a los egresados con el fin de que logren luego de postularse a las ofertas laborales superar los procesos de selección de las empresas, se plantea mediante la ejecución del presente trabajo dotar a las personas que se encargan de moderar la interacción entre empresas y egresados entorno a las ofertas laborales, con un espacio web para acceder a recomendaciones de competencias académicas obtenidas mediante consultas a un repositorio RDF obtenido de la transformación de datos relacionales con que se cuenta.

Figura 1-3: Comunidad Red de Prestadores de Servicio Público de Empleo.



Fuente: <http://comunidad.serviciodeempleo.gov.co>

Finalmente se definirá una consulta que permita realizar búsquedas con temáticas específicas que las empresas requieran para cubrir sus vacantes.

1.3. Objetivos

Objetivo General: Desarrollar un sistema que permita obtener información para orientar en la selección de estudios complementarios adicionales a profesionales para recién egresados que puedan mejorar su empleabilidad, mediante la estructuración jerárquica de términos u ontologías OWL (Lenguaje de Ontologías Web), SWRL (Lenguaje de Reglas de la Web Semántica) y NLP (Procesamiento de Lenguaje Natural). A continuación, se mencionan los objetivos específicos:

- Transformación de datos los datos almacenados en bases de datos de modelo Entidad Relación ER a Esquema de Descripción de Recursos RDF.
- Procesamiento de datos en lenguaje natural presentes en los perfiles de los currículos de los egresados recientes.
- Integrar al modelo de consultas una ontología para obtener las tripletas recomendadas los términos de la base de perfiles demandados en las ofertas laborales que el mercado laboral demanda para los perfiles.
- Obtención y almacenamiento de las inferencias o reglas, a partir del conocimiento y las competencias de los egresados.
- Proporcionar al usuario una interfaz de interacción que permita a los recién egresados a través los coordinadores de la bolsa de empleo para obtener un listado de posibles estudios que mejoren su perfil de empleabilidad.

1.4. Alcances

El alcance está delimitado por la cantidad y calidad de los perfiles de los egresados que se gradúan en la universidad y que se registran en la base de datos del SIE para los 80 programas de pregrado, 293 programas de posgrado. Del mismo modo se contempla como limitante las ofertas demandadas por las empresas que mediante su registro y selección permitirán identificar los perfiles que destacan según su programa de egreso.

Ante la dificultad en la obtención de primer empleo por parte de recién egresados, las bolsas laborales orientan a los usuarios en la manera en que se postulan a las ofertas

demandadas, por lo que la bolsa de empleo de la Universidad Nacional de Colombia cuenta con gestión de ofertas en las sedes ubicadas a lo largo del territorio nacional. El alcance de los productos del presente trabajo solo estará disponible para la oficina de nivel nacional del programa de egresados que opera desde Bogotá.

El presente proyecto forma parte de la exploración de nuevos servicios que el programa revisa para complementar los servicios de la bolsa de empleo, por lo cual solo estará disponible para orientar a los usuarios una vez sean surtidos los procesos de validación que la institución estime pertinentes en consonancia con las normas internas, locales y nacionales.

1.5. Limitaciones

Una de las limitaciones más notorias se presenta al momento de realizar la extracción de etiquetas o anotaciones en lenguaje natural, debido que los motores disponibles actualmente cuentan con modelos de entrenamiento mayormente desplegados en el idioma inglés, y a pesar de que algunas de las herramientas permiten el uso de complementos o métodos en sus APIs para habilitar idiomas adicionales como el español para el presente caso, los resultados derivados de las pruebas muestran un desempeño más pobre en anotación semántica automática, por lo que las etiquetas presentadas en la interfaz serán resultado del procesamiento sugerencias en idioma inglés. De otro lado el sistema no contempla un método que valide la vigencia de las recomendaciones, es decir los tiempos adecuados para que las recomendaciones sean consideradas como efectivas para lograr el impacto deseado en la favorabilidad de la selección de los currículos de los graduados por parte de las empresas.

1.6. Metodología

- Obtención del acceso a la base de datos de egresados de manera segura, y que atienda formalmente a la ley de reserva estadística, de modo que se impida que los datos no podrán darse a conocer al público ni a las entidades u organismos ajenos a la Universidad, sino únicamente en resúmenes o resultados de agrupaciones, que no hagan posible

deducir de ellos información alguna de carácter individual que pudiera utilizarse para fines comerciales, de tributación fiscal, de investigación judicial o cualquier otro diferente del propiamente estadístico. Ley de Reserva Estadística (Art. 5 Ley 79 de 1993).

- Se realiza una revisión de las soluciones informáticas disponibles de uso limitado o gratuito que permitan llevar a cabo el procesamiento sobre los perfiles de hojas de vida, así como las de soluciones informáticas libres que soporten escalamientos futuros.
- Con el fin de no afectar la operación de la Base de datos de producción se diseñan y optimizan las consultas que permitan obtener un reporte de los perfiles de egresados.
- Selección de una herramienta ETL que permite preparar la información de los perfiles de los profesionales demandados y disponerla en el formato específico requerido para los procesamientos en cada una de las etapas del proyecto.
- Selección de una herramienta que permita almacenar y generar consultas SPARQL sobre los términos obtenidos.
- Selección de una herramienta que permita generar las interfaces de consulta.

2. Marco Teórico

Se acepta en el país y en los sistemas de educación superior del mundo que los egresados, continúan durante toda su vida, siendo una parte activa de las comunidades universitarias de las Instituciones de Educación Superior de donde se han graduado; por esto desde la promulgación de la Ley 30 de 1993 y posteriormente con el Sistema de Acreditación de la Educación Superior colombiano, se promueve tanto el seguimiento a los egresados para conocer la pertinencia de la formación, como el involucramiento en la vida universitaria.

Los egresados se constituyen en un factor clave de éxito en cuanto a la evaluación de la calidad de los procesos académicos y de los impactos esperados y logrados en la sociedad.

2.1 Calidad en la educación y empleabilidad

El concepto de Calidad ha sido un referente permanente en las acciones del ser humano. La calidad es un concepto complejo, por lo que se comprende mejor desde una perspectiva sistémica. La calidad es un concepto dinámico relacionado con la bondad de un bien o servicio, en términos del valor agregado, que cada usuario perciba en él; es decir, que tal objetivo se otorga a los productos dotados del conjunto de atributos deseados para satisfacer las necesidades y expectativas de los destinatarios de los esfuerzos de la organización.

El enfoque de estructura procesos y resultados se emplea para caracterizar y analizar la calidad, con la cual se adelantan las funciones básicas de las Instituciones Educativas. Se

entiende por estructura, todas las características relativamente estables de las entidades de educación superior e incluye los recursos humanos, físicos y financieros y la forma como estos se organizan; abarcando características como el número de calificación y distribución del personal, el tipo de práctica (individual o grupal), las formas de remuneración, la presencia de auditoría, equipos, suministros y demás recursos.

Por otra parte, Procesos, son el grupo de actividades que llevan a cabo internamente y externamente las universidades, internamente se realizan actividades por y entre profesores administrativos y estudiantes, que constituyen el proceso de educación, encontrándose tres clases: los procesos puramente académicos, los procesos administrativos académicos, y los procesos administrativos. Externamente se encuentran las diferentes relaciones que establecen con otras instituciones, estos procesos constituyen el funcionamiento de una universidad.

Finalmente, los resultados, son los cambios del estado actual y futuro del nivel de educación de un estudiante, que puede ser atribuido al antecedente de la formación; incluyendo un mejoramiento de las competencias y cambios de comportamiento relacionados con el nivel de formación (Programa de Egresados Universidad Nacional de Colombia, 2009). Recientemente se habla de impacto referido a los efectos colectivos asociados a la acción educativa dentro de los cuáles se analizan los cambios en los patrones de especialización de los países. Tales evaluaciones se han adelantado con la participación de los egresados.

El sistema de calidad consta de tres elementos, los cuales son responsabilidad de la unidad administrativa encargada, estos son: 1) Procedimientos a nivel organizacional, 2) Control de calidad enfocado a tareas específicas y 3) Garantía de calidad.

La garantía de calidad es una actividad de la gestión de la calidad, la cual se inicia con la construcción de una política de la calidad que establece objetivos y estrategias; dentro de estas últimas se encuentran: el sistema de calidad, la organización de la calidad y la garantía de la calidad. La política de calidad debe establecer claramente el compromiso de alta dirección con los principios de calidad; los cuales deben ser difundidos, comprendidos e implementados garantizándose su ejecución.

Por otro lado, los objetivos de la calidad deben considerar la capacidad de la organización para desarrollar los procesos, servicios y los sistemas necesarios para la política de calidad. Un sistema de calidad integra tanto el análisis de las necesidades y expectativas de los distintos grupos objetivo, como la planeación del desarrollo de nuevos servicios, compras, inspección, así como un sistema de auditoría de la calidad.

La garantía de la calidad es un proceso sistemático de verificación que certifica que la inspección de la calidad y las operaciones de control de calidad se realicen en la forma debida; al igual que evalúa el trabajo interfuncional orientado a mantener y mejorar el nivel de calidad definido. Esta actividad debe proveer a las partes interesadas evidencia que genere confianza sobre la función calidad.

La garantía de calidad tiene como estrategias: 1.) el planeamiento de la calidad, 2.) el control de la calidad 3.) y la auditoría de la calidad. Dentro del planeamiento se consideran las metas de la calidad a alcanzar, el sistema que garantiza la calidad en cada etapa de los procesos de la organización, la elección de la tecnología, materiales, equipos y procesos necesarios, definición de competencias de los empleados; 2.) el control busca establecer si el nivel de calidad planeado y las estrategias formuladas para alcanzarlos se han cumplido en la ejecución. 3.) La auditoría de la calidad verifica periódicamente si las actividades necesarias para alcanzar los objetivos de la calidad están siendo implementadas, si los resultados alcanzados están dentro de los niveles planeados.

La presión general sobre la calidad en la oferta de bienes y la prestación de servicios ha generado una amplia difusión de iniciativas que buscan determinar, mantener y mejorar los niveles de desempeño de las organizaciones, tal situación ha conducido a la evolución gradual de las estrategias que propenden por garantizar la calidad. Se pueden caracterizar entonces varias etapas que representan niveles cualitativos distintos y los cuales se construyen mediante un proceso acumulativo, desde el Licenciamiento por inspección; la Certificación de procesos y programas; La acreditación de programas e instituciones y Adopción de Modelos de Excelencia para sedes e instituciones.

Respondiendo a lo anterior, en diciembre del 2003 se expidió la Ley 872 por la cual se crea el sistema de gestión de la calidad en la Rama Ejecutiva del Poder Público y en otras

organizaciones prestadoras de servicios, como herramienta de gestión sistemática y transparente que permita dirigir y evaluar el desempeño institucional, en términos de calidad y satisfacción social en la prestación de los servicios a cargo de las organizaciones y agentes obligados, la cual estará enmarcada en los planes estratégicos y de desarrollo de tales organizaciones. El sistema de gestión de la calidad debe adoptar en cada organización un enfoque basado en los procesos.

Esta ley prevé sistemas de Certificación y de Acreditación para las organizaciones públicas los cuales deben ser reglamentados.

Garantía de la Calidad basada en la Acreditación es a menudo un proceso voluntario en el cual las organizaciones deciden participar. En este una organización reconocida, generalmente una organización no gubernamental (ONG), valora y reconoce que un área u organización cumple con los estándares publicados, aplicables y preestablecidos. Los estándares de acreditación generalmente se consideran óptimos y factibles y han sido diseñados para promover el mejoramiento continuo dentro de las organizaciones. La decisión acerca de la acreditación de una organización específica se toma después de una evaluación realizada por un equipo y usualmente se hace cada dos o tres años. Frente a los Procesos de Acreditación el sector Universitario ha sido pionero y desde 1994 cuenta con el Decreto 2904 que establece el Sistema Nacional de Acreditación; en Salud el Decreto 2309 de 2002 y la Resolución 1474 de 2002 estructuran el Sistema Nacional de Acreditación en Salud.

Garantía de la calidad del sistema de educación superior La Calidad es un concepto multidimensional el cual puede analizarse desde los insumos o desde los productos; desde los insumos la calidad, resulta de la disponibilidad, del uso y de la combinación de los recursos. Por otro lado, desde los productos la calidad existe cuando se cumplen las metas en relación con ciertos estándares. Pero también se usa otro enfoque, el de los procesos, en donde estos tienen relación con aspectos de organización y gerencia. Desde otra perspectiva, el concepto de calidad en la educación superior hace referencia a un atributo del servicio público de educación y el modo como tal servicio se presta. La ingeniería es una profesión sensible a los cambios sociotécnicos, sus ámbitos de aplicación son: la producción, la gestión e investigación y desarrollo; tiene como recurso básico al a tecnología, por esto la crea la modifica y la aplica.

La calidad se vincula al conjunto de atributos, elementos, procesos y funciones que permiten acercar un programa de ingeniería a los objetivos del programa. Mientras la pertenencia se asocia a la concordancia entre los objetivos y las expectativas de quienes impulsan o promueven dichos programas; Toda vez que un programa de ingeniería es una organización que ofrece procesos educativos capaces de formar profesionales con las características demandadas por cada área.

En general, la calidad de una profesión se reconoce por la articulación y la coherencia, entre el campo de conocimiento y el campo de práctica, como por la demostración de las competencias necesarias. El gobierno Nacional ha impulsado la construcción participativa de un conjunto de estándares de calidad entendidos como niveles o grados necesarios e indispensables para que algo pueda ser aceptable, Entre los elementos de los estándares se tienen; los referentes, las dimensiones y las prácticas. Los estándares se establecen respondiendo a las dimensiones y las prácticas, para contribuir o definir la existencia y la identidad, en correspondencia con las necesidades y características de cada programa.

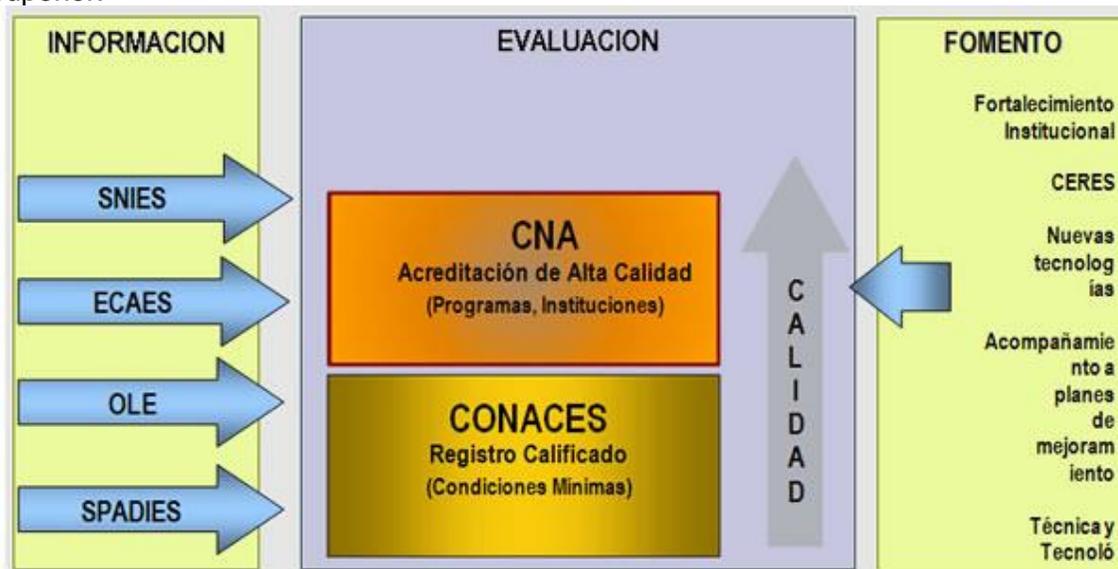
En Colombia, la ley 30 de 1992 establece en el artículo 31: “propender por la creación de mecanismos de evaluación de la calidad de los programas académicos de las instituciones de educación superior “; tal norma define dos mecanismos básicos: 1) Acreditación y 2) exámenes de estado a los estudiantes que egresan de la educación secundaria.

El sistema se materializa en 1995 cuando se implementa el Sistema Nacional de Acreditación y a partir del 2000 se ha difundido paralelo con un cambio de enfoque, a uno centrado en las competencias. Por otro lado, en el 2003 se modificó la estructura organizativa del Sistema de Aseguramiento de la Calidad; se creó el Viceministerio de Educación Superior y la subdirección de Aseguramiento de la Calidad.

En el proceso de Acreditación de educación superior se distinguen dos aspectos: el primero es la evaluación de la calidad realizada por la institución misma; también como por agentes externos que pueden entender la naturaleza de lo que se evalúa y por el Consejo nacional de Acreditación; el segundo es el reconocimiento público de la calidad realizado por el CNA. Existen tres componentes que se traducen en tres etapas del proceso de acreditación. Son ellas: 1) La auto evaluación, que consiste en el estudio que llevan a cabo

las instituciones o programas académicos, sobre la base de los criterios, características e indicadores, definidos por el Consejo Nacional de Acreditación. 2) La evaluación externa o evaluación por pares, utiliza la auto evaluación y verifica sus resultados, y concluye en un juicio sobre la calidad. 3) La evaluación final que realiza el Consejo Nacional De Acreditación. Y la cual se orienta al reconocimiento de la calidad por el cual se da fe pública.

Figura 2-1: Arquitectura del Sistema de Aseguramiento de la Calidad de la Educación Superior.



Fuente: Consejo Nacional de Acreditación

El gobierno nacional ha creado un Sistema de Aseguramiento de la Calidad de la Educación Superior, conformado por los organismos, las acciones y las estrategias que aplican desde el proceso mismo de creación y establecimiento de una institución de educación superior, hasta el desempeño del profesional que egresa del sistema (Consejo Nacional de Acreditación, 2004). La arquitectura del mismo se presenta en el siguiente diagrama. Adicionalmente también algunas universidades certifican sus procesos a través de las normas ISO 9000, al igual que otras certifican sus programas empleando organismos internacionales.

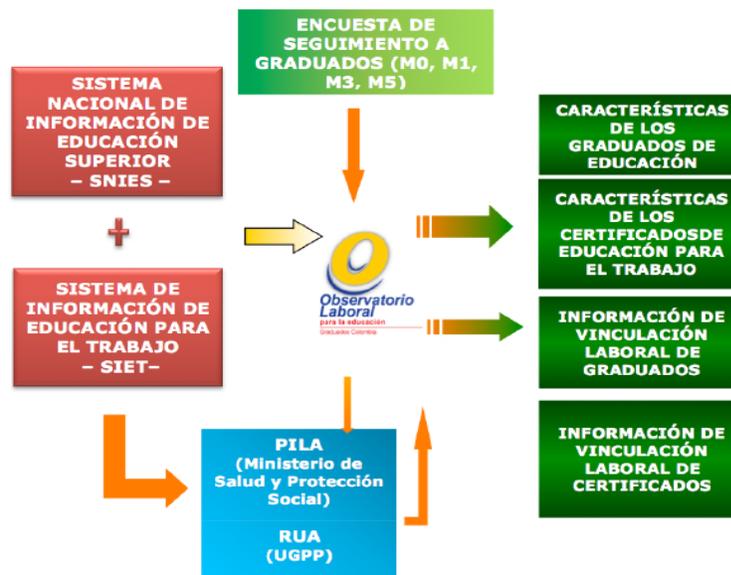
La Ley 30 de 1992 da a las instituciones de educación superior la libertad de acogerse al Sistema Nacional de Acreditación, lo cual, aunque no reemplaza la inspección y vigilancia

que ejerce el Estado, introduce un énfasis de calidad que ha logrado comprometer a dichas instituciones con sus egresados como resultado misional de dichas instituciones.

Antes de esta ley el problema de las universidades estaba centrado en el establecimiento de los conjuntos de materias que habrían de conformar los currículos de los programas académicos y en los términos en que se debían cumplir con dichos contenidos, lo cual determinaba qué y cómo se debe estudiar para obtener un título específico. A pesar de esto, no se hacía tanto énfasis en la calidad específica de los contenidos de tipo práctico ya que eran un complemento a los contenidos teóricos, de tal modo que la cantidad de horas impartidas en los laboratorios no siempre contaban con los elementos necesarios para lograr la calidad deseable en el entorno laboral, lo cual le ocasionaba a los egresados vacíos ocupacionales a la hora de desempeñarse laboralmente.

Lo anterior llevó a las IES a responder preguntas relacionadas con el desempeño de sus egresados como por ejemplo si consiguen o no empleo y cómo los remunera el sector laboral. De este modo las universidades han empezado a enfocarse en el egresado como su finalidad y resultado y no la enseñanza exclusivamente. Es decir, el egresado como producto del ejercicio institucional de las IES formado para el primer empleo profesional.

Figura 2-2: Entradas y salidas del Sistema O3.



Fuente: Ministerio de Educación Nacional - Observatorio Laboral para la Educación

Actualmente el desempeño de los egresados puede ser monitoreado de manera sistémica gracias a herramientas como el O3 del ministerio de Educación nacional que permite consultar tanto la vinculación laboral de los recién graduados como la de la totalidad de los egresados, llegando al detalle de los montos devengados en el mercado laboral gracias a que dicho sistema se abastece de datos de cotización y de Ingreso Base de Cotización promedios del año de estudio junto con los Datos de graduados de instituciones de educación superior.

El observatorio laboral para la educación genera análisis anuales lo cual favorece la inspección del desempeño de los profesionales en el ámbito laboral, sin embargo los resultados de estos análisis no serán recomendables para los aspirantes a ingresar a un programa académico en particular, debido a que no hay certezas en que esta información usada para determinar cuáles son los programas más demandados hoy, corresponderán con los datos al final de su formación, por lo que este observado desfase siempre está presente en los ciclos de formación. De modo similar los resultados en un momento dado por este tipo de sistemas como lo es el propuesto en el presente trabajo, tampoco son recomendables para generar cambios curriculares debido a que dichos cambios normalmente son dados por cambios de la economía, por el auge sostenido en un sector de la economía en particular o por cambios tecnológicos que evidencian un rezago en capacidades con que cuenta los profesionales actuales impidiendo que cuenten con las semejanzas o particularidades específicas requeridas.

De lo anterior se observa para el presente trabajo la necesidad de establecer una periodicidad en la consolidación de la información que responda de manera viable al objeto de estudio, con el fin de generar resultados que entendidos como recomendaciones puedan ser sugeridas de manera oportuna, en la sección 5.2 se dará una periodicidad inicial con la que se procure dar continuidad al procesamiento de la información unos resultados con la vigencia más adecuada.

En general se observa que instrumentos como el O3 del Ministerio de Trabajo brinda insumos al sistema de educativo actual que le permite evaluar de manera retrospectiva el desempeño de los recién egresados, con el presente trabajo se pretende dar elementos a los recién egresados que les permita desde lo individual realimentar y enriquecer su currículo con temáticas y estudios complementarios entendidos como recomendaciones

dado que son obtenidos con base en los perfiles que son seleccionados por las empresas por cumplir con lo necesario para satisfacer las exigencias de las ofertas laborales. A continuación, se presentan conceptos básicos adicionales que pueden ser de ayuda para entender el desarrollo de proyecto desde su nivel estructural o general hasta su componente tecnológico.

2.2 Currículo

Elemento que permite a los individuos plasmar formalmente su trayectoria académica laboral y en general agrupar los componentes tanto de aptitudes como de actitudes y que es definido por el ministerio de educación como “El conjunto de criterios, planes de estudio, programas, metodologías, y procesos que contribuyen a la formación integral y a la construcción de la identidad cultural nacional, regional y local, incluyendo también los recursos humanos, académicos y físicos para poner en práctica las políticas y llevar a cabo el proyecto educativo institucional” (Ministerio de Educación Nacional, 2013).

2.2.1. Perfil profesional

En años recientes han avanzado iniciativas como el Proyecto Tuning (Beneitone et al., 2004) y Tuning Latin America (Beneitone, González, & Wagenaar, 2014) las cuales han permitido llevar a cabo reflexiones a nivel internacional en torno al conjunto de competencias que conforman el perfil profesional. Dichas competencias entendidas de modo general en el ámbito nacional como el “Conjunto de conocimientos, actitudes, disposiciones y habilidades (cognitivas, socio-afectivas y comunicativas), relacionadas entre sí para facilitar el desempeño flexible y con sentido de una actividad en contextos relativamente nuevos y retadores. Por lo tanto, la competencia implica conocer, ser y saber hacer.”(Ministerio de Educación Nacional, 2013). Dicho conjunto de conocimientos, habilidades y actitudes que aplicadas y/o demostradas en situaciones del ámbito productivo, tanto en un empleo como en una unidad para la generación de empleo por cuenta propia, se traducen en resultados efectivos que contribuyen al logro de los objetivos de la organización o negocio. Las competencias laborales pueden ser generales o específicas.

2.2.1.1. Competencias laborales generales

Son las competencias requeridas para desempeñarse en cualquier entorno social y productivo, sin importar el sector económico, el nivel del cargo o el tipo de actividad, pues tienen el carácter de ser transferibles y genéricas. Además, pueden ser desarrolladas desde la educación básica primaria y secundaria, y por su carácter, pueden coayudar en el proceso de formación de la educación superior y, una vez terminado este ciclo, a una vida profesional exitosa.

2.2.1.2. Competencias laborales específicas

Son las competencias que están orientadas a habilitar a un individuo para desarrollar funciones productivas propias de una ocupación o funciones comunes a un conjunto de ocupaciones. En el sector educativo, estas competencias han sido ofrecidas tradicionalmente por las instituciones de educación media técnica.

2.3 Mercado de trabajo profesional

Partiendo de una definición del mercado como “el lugar en el que se encuentran los que buscan empleo (demandantes) y los que ofrecen trabajo (empleadores)” (Francisco Burgos Becerra, 2002). Donde en general, los empleadores son organizaciones públicas o privadas, y los demandantes de empleo son aquellas personas que cumplen con la formación de nivel profesional requerida por la oferta laboral.

Una reciente investigación conducida por la Unidad de evaluación de Políticas del departamento de Investigaciones de la Organización Internacional de Trabajo OIT denominado “Soluciones Eficaces: Políticas activas del mercado de trabajo en América Latina y el Caribe” muestra como el mercado de trabajo tiene una especial relevancia a nivel estratégico para hacer frente al estancamiento de la creación de empleos sostenibles y de calidad y como las políticas implementadas en esta materia pueden mejorar las

condiciones de vida de los trabajadores y de sus familias e, incluso, mejorar la equidad al contribuir en la reducción de la pobreza.

Recientemente el mercado laboral colombiano ha sido objeto de la intervención gubernamental mediante el Decreto 722 de 2013 del Ministerio del Trabajo que reglamenta la prestación del Servicio Público de Empleo, la actividad de intermediación laboral y la conformación de la red de operadores del Servicio Público de Empleo, conforme a lo previsto en el artículo 25 de la Constitución Política, en que se declara al trabajo como un derecho que requiere especial protección del Estado.

2.4 Necesidades del mercado laboral

En Colombia el Observatorio laboral publicó un análisis de información entre el 2001 y el 2013 donde se evidencia que el 8,4% de los graduados de la educación superior (títulos otorgados por Instituciones de Educación Superior), proviene de programas asociados al sector de Tecnologías de la Información y Comunicación (TIC). De los 229.197 graduados vinculados a este sector, el 47,9% son del nivel universitario, el 31,2% son Tecnólogos y el 6,8% son Técnicos Profesionales, lo que muestra un incremento en la acogida que están teniendo los programas de estos dos últimos niveles de formación, algo significativo cuando una de las metas es fortalecer la educación Técnica Profesional y la Tecnológica (Observatorio Laboral para la Educación, 2013). Al sector de las TIC, le siguen en número de graduados los relacionados con el sector minero energético (111.468) y el sector agroindustrial (107.103 titulados). Para ese momento se destacó, sin embargo, que programas de educación superior asociados a otros sectores, agrupa el 77,5% de los graduados de todo el país.

Particularmente, en 16 de los 20 sectores establecidos por el Programa de Transformación Productiva (PTP), los de mayor concentración de graduados están en el sector BPO (21,5%), Software y TIC (7,3%), Palma y Combustibles (2,5%) y Chocolatería (2,5%). En total, estos 16 sectores concentran el 45,8% de los egresados en carreras asociadas a los mismos. Por otro lado, al ver el comportamiento de aquellos que estudiaron programas de educación para el trabajo y el desarrollo humano entre el 2010 y el 2013, el sector de

Agroindustria tiene el 15,1% de los egresados, seguido del de TIC con el 6,2% y el de Vivienda con el 4,5%.

2.5 Desajustes del mercado laboral entre la oferta y la demanda

Un estudio elaborado por la Fundación Europea Sociedad y Educación, encontró para uno de los países miembro de la unión europea, un gran porcentaje de trabajadores que al ser encuestado consideró que para desempeñar su puesto de trabajo no sería necesario tener un título universitario, es decir, que consideran estar sobrecualificados para el puesto actual de trabajo. Comparativamente en dicho estudio España es uno de los países con mayor nivel de desajuste educativo, tanto en términos de cualificaciones como de competencias. Según datos procedentes de la OCDE, el 22% de los trabajadores españoles tiene más educación de la que sería requerida por su puesto de trabajo (over-qualification), sólo superado por Reino Unido, Australia, Irlanda y Estonia. Además, España es, sólo superado por Austria, el país donde una proporción más alta de trabajadores tienen más competencias de las que se necesitarían para el puesto (over-skilled).

En Colombia el observatorio del mercado de trabajo y la seguridad social de la Universidad Externado de Colombia analizó recientemente el desajuste del mercado laboral encontrando en 2016 que un 30,3% de los ocupados tiene un nivel educativo que no se ajusta a los requerimientos que sus oficios exigen de los cuales mientras que un 20,1% esta sobrecualificada, lo que muestra que solo el 49,7% cuenta con un ajuste adecuado a su nivel de cualificación (Universidad Externado de Colombia, 2017). Con lo cual se tiene que el 30,3% de las personas no suficientemente cualificadas requerirán estudios adicionales y que solo la pertinencia en dichos estudios afectará en forma positiva estas cifras.

2.6 Pertinencia en la formación

La Conferencia Mundial sobre la Educación Superior, señaló que la educación debe tratar de alcanzar simultáneamente los objetivos de equidad, pertinencia y calidad, situando a la pertinencia como elemento necesario para cumplir su cometido con autonomía, eficacia,

transparencia y de cara a la responsabilidad social (UNESCO, 2009). En análisis empíricos del mercado del trabajo, la Cepal encuentra que la relación de la profesión con la ocupación es múltiple mediante encuestas que miden las personas ocupadas con datos no solo de la oferta o demanda, sino el punto donde confluyen los trabajadores que ofrecen su trabajo a un salario y la demanda de los empleadores. (Sarmiento & Baldión, 2014).

Figura 2-3: Elementos básicos del Sistema Nacional de Formación Técnica



Fuente: El Desafío de Aumentar la Pertinencia de la Formación Profesional. (Sarmiento & Baldión, 2014)

La Pertinencia y la calidad de la formación dependen de la inserción en los sistemas social, económico y de participación política. Pero para lograrlo es necesario no solo que se dé una adecuada formación, sino que se desarrollen subsistemas de apoyo. En el marco normativo colombiano se está buscando en la actualidad construir un sistema que permita que las personas puedan formarse por diferentes vías, complementarias entre sí, a continuación, se muestra un diagrama con los elementos básicos del Sistema Nacional de Formación Técnica, según el Consejo Nacional de Política Económica y Social (CONPES). De esta manera la formación académica, la formación profesional y aún la experiencia son rutas posibles de formación y calificación. Sin embargo, deben ser reconocidos por el mercado para los cual se requieren lo sistemas de apoyo mencionados.

En la actualidad además de la formación en los aspectos cognitivos necesarios para desempeñarse profesionalmente, se requiere lograr el desarrollo integral de la persona lo que implica aprender a trabajar en equipo, poder actuar de manera autónoma, y de manera

integrada con equipo heterogéneos. Se trata de contar con competencias claves que permiten identificar y resolver problemas complejos de modo colectivo.

2.7 Formación Continua

Es un modelo que se articula con otros sistemas de organización productiva y de actuación social. Es un sistema de calidad que permite definir, evaluar y desarrollar las competencias (cualificaciones) de las personas en relación con los resultados que deben lograr en el trabajo en general o en un empleo en particular (Carvallo, Aymamí, 2013). Universidades como la Nacional de Colombia cuyos ejercicios de autonomía le permite a nivel de su consejo superior Universitario definir la educación continua y permanente como la modalidad de Extensión que se realiza mediante cursos de extensión, actualización o profundización, diplomados y programas de formación docente, articulados con los programas académicos de pregrado y postgrado de la Universidad.

El Sistema que más ha sido difundido es el del Reino Unido que se fundamenta en un proceso de normalización, en el que se pone en común el lenguaje que permitirá el diálogo entre el mundo del trabajo y el de la formación, mediante la definición o actualización de los estándares que describen los conocimientos, habilidades y destrezas requeridos para el desempeño de funciones productivas; Al relacionar y normalizar los requerimientos de competencias de los sectores productivos, estas sirven de referente para el desarrollo curricular. La formación basada en competencias permite avanzar progresivamente en la adquisición de niveles de cualificación cada vez más amplios.

El proceso de evaluación y certificación, que permite validar que una persona ha alcanzado la calidad de desempeño esperado y el nivel de competencias descritas en las normas definidas previamente para una función específica. La evaluación de dichas competencias se realiza por instituciones acreditadas por un organismo certificador, bajo las políticas y normas de la autoridad reguladora, que constata y certifica que las personas son competentes. Por último, el monitoreo de la Información sobre capital humano y de los principales resultados da insumos para evaluar y mejorar las políticas públicas. Los datos provienen de distintas fuentes como censos, encuestas de hogares y/o de empresas,

registros administrativos, con preguntas específicas sobre formación, sobre la oferta y la demanda de capital humano y la formación de este.

2.8 Web Semántica

Es un conjunto de estándares y tecnologías útiles en el propósito de generar una Web de Datos (Yu, 2011), esto significa que se hará uso de Datos que estarán Enlazados por metadatos que le dan sentido o significado generados a partir de dicho conjunto de elementos constitutivos que junto con la Web Semántica, generaría una Web extendida de Datos enlazados donde estos últimos se publican con las tecnologías de Web Semántica y atendiendo a las mejores prácticas y estándares respectivos, de otro modo se puede decir que La Web semántica es una finalidad, donde los Datos Enlazados proporciona los medios para llegar a dicho fin.

3 Estado del arte

Al revisar experiencias similares de sistemas que cuentan con elementos de Web semántica para realizar búsquedas sobre conjuntos de datos en formato RDFs, se encuentran componentes, estándares y conceptos comunes que pueden ser revisados a partir de la categorización realizada en (Tzitzikas, Manolis, & Papadakos, 2017) donde se abordan sistemas para explorar conjuntos de datos de tipo de entidad única como también sistemas para explorar conjuntos de datos de tipo de entidades múltiples, a continuación se muestra una comparativa de estos sistema respecto a dichas fuentes de datos junto con el detalle de sus salidas y el espacio de información en que se da su procesamiento.

Tabla 3-1 Comparativo de sistemas y sus fuentes RDF para experiencias similares adaptado de Tzitzikas, Manolis, & Papadakos, 2017.

Sistema	Salidas	Espacio de Información	Fuentes de Datos
BrowseRdf (Oren, Delbru, & Decker, 2006)	Registros	Orientado a Objetos	Única
Camelis2 (Ferré, 2010)	Conteo	Orientado a Objetos	Única
/Facet (Hildebrand, Wiskunde, Ossenbruggen, & Wiskunde, 2006)	Conteo	Orientado a Objetos	Única
Faceted data explorer (Erling, 2009)	Conteo	Orientado a Objetos	Única
Faceted Wikipedia (Hahn et al., 2010)	Conteo	Orientado a Objetos	Única
Fuzzy view based search (Holi & Hyvönen, 2006)	Registros	FRDF	Única
gFacet (Heim, Ertl, & Ziegler, 2010)	Registros	Orientado a Objetos	Única
GRQL (Athanasios, Christophides, & Kotzinos, 2004)	Registros	Orientado a Objetos	Única
Humboldt (Kobilarov & Dickinson, 2008)	Registros	Orientado a Objetos	Única
Longwell (Pietriga, 2006)	Conteo	Orientado a Objetos	Única
MediaFaces (van Zwol et al., 2010)	Conteo	Orientado a Objetos	múltiple
mSpace (schraefel et al., 2005)	Registros	Orientado a Objetos	Única

MuseumFinland (Hyvönen et al., 2004)	Conteo	Orientado a Objetos	múltiple
Odalisque (Allard et al., 2008)	Conteo	OWL	Única
Ontogator (Mäkelä et al. 2006)	Conteo	Orientado a Objetos	Única
Parallax (Huynh & Karger, 2009)	Conteo	Orientado a Objetos	múltiple
Rhizomer (Brunetti, Garcia, & Auer, 2013)	Conteo	Orientado a Objetos	Única
SemFacet (Arenas, Cuenca Grau, Kharlamov, Marcuska, & Zheleznyakov, 2014)	Registros	OWL 2	Única
SPARKLIS (Ferré, 2015)	Conteo	Orientado a Objetos	Única
VisiNav (Harth, 2009)	Conteo	Orientado a Objetos	múltiple

La revisión de los sistemas incluidos como experiencias similares en cuanto a sistemas de exploración de conjuntos de datos en formato RDF se logra obtener una relación de elementos comunes que a continuación permite detallar las características asociadas a cada uno de los componentes, conceptos y estándares que habilitan las prestaciones brindadas.

Tabla 3-2 Resumen de componentes, estándares y conceptos centrales del estado del arte.

Elemento	Tipo	Características
XML	Estandar	Generalmente usado como formato que facilita el procesamiento y transmisión de los datos.
XSD	Estandar	Generalización del xml cuyo propósito es validar la estructura de otro documento xml. El xsd no es obligatorio para ningún xml, pero asegura que el xml podría usarse para algunos propósitos particulares. El xml solo contiene datos en formato y estructura adecuados.
Ontologías o Vocabularios	Conceptos	Las ontologías y los vocabularios están relacionadas a la forma en la cual, sintácticamente, los metadatos pueden agregarse a las informaciones con el objetivo de conseguir que el significado pretendido por el publicador de los datos sea el mismo que el significado entendido por el consumidor de los mismos,
RDF (Resource Description Framework)	Estandar	Es una estructura para representar informaciones en la Web. El RDF permite realizar afirmaciones sobre recursos. Cualquier cosa puede ser denominada recurso, tanto cosas concretas como abstractas. Una empresa determinada, una persona, una página Web, constituyen recursos. Un sentimiento, o un color, también son considerados recursos.
SKOS	Estandar	Sintaxis que ofrece una manera de representar esquemas de clasificación, tales como vocabularios controlados, taxonomías y tesauros. En SKOS, los conceptos puede ser identificados mediante URIs, con rótulos en una o más lenguas, y ser documentados con diferentes tipos de notas.
OWL Lenguajes de	Estandar	Es un lenguaje que extiende RDF y RDFS, ofreciendo un conjunto mucho más amplio de tipos de restricciones al

Ontologías Web		conjunto de tripletas definidas. Además de ello, el lenguaje ofrece una serie de diversos constructores que permiten, entre otras posibilidades, la construcción de clases complejas a partir de otras definiciones de clases, y el encadenamiento de propiedades.
RIF Formato de intercambio de reglas	Estandar	Iniciativa que procura un estándar para el intercambio de reglas entre los sistemas de reglas, en particular entre los motores de reglas web. El medio de intercambio entre diferentes sistemas de reglas es XML, un formato para el intercambio de datos.
Editor de Ontologías Protege	Componente	Herramienta para el modelado y edición de conocimiento, es un framework de código abierto escrita en Java para el cual otros proyectos sugieren plugins.
Datos Enlazados	Estandar	Establece el acceso a un recurso a través de su URI, donde deben devolverse informaciones relevantes sobre el recurso identificado (desreferenciación).
Contenedores de RDF	Componente	Sistemas de Gestión de Bases de Datos que admiten almacenamiento de datos en formato de tripletas RDF.
Lenguajes de Consulta	Estandar	Es un lenguaje de computadora, específicamente un lenguaje de consulta para bases de datos, capaz de recuperar y manipular datos almacenados en el formato RDF.
SPARQL	Estandar	SPARQL se ha convertido en el lenguaje de consulta RDF estándar y en 2008 se convirtió en una recomendación del W3C. Los lenguajes de consulta RDF se pueden agrupar en familias de idiomas. La familia de lenguajes SPARQL incluye SquishQL, RDQL, SPARQL y TriQL. Estos lenguajes tratan los almacenes de datos RDF como almacenes de tripletas que no necesariamente tienen información ontológica o de esquema asociada a ellos.
End-point SPARQL, Punto final de acceso	Componente	Servicio de protocolo SPARQL que permite a los usuarios consultar una base de conocimiento a través del lenguaje SPARQL.

A continuación, se detallan los estándares, conceptos y componentes resumidos en la tabla 3.2.

3.1 Estándares

3.1.1 XML

El lenguaje de marcado extensible es un lenguaje sencillo de etiquetas que ofrece un formato portable de documentos ya que se puede afirmar que es multiplataforma, y gracias a que puede representar cualquier tipo de dato que se pueda codificar como texto se flexible porque puede representar [Harold et al; 2005]. Se compone de etiquetas de

apertura "<nombre_tag>" y de una correspondiente etiqueta de cierre "</nombre_tag>" lo que se encuentre entre ambas es el valor del elemento.

3.1.2 XSD

El lenguaje de marcado extensible es un lenguaje sencillo de etiquetas que ofrece un formato portable de documentos ya que se puede afirmar que es multiplataforma, y gracias a que puede representar cualquier tipo de dato que se pueda codificar como texto se flexible porque puede representar [Harold et al; 2005]. Se compone de etiquetas de apertura "<nombre_tag>" y de una correspondiente etiqueta de cierre "</nombre_tag>" lo que se encuentre entre ambas es el valor del elemento.

3.1.3 RDF / RDFS

RDF (Esquema de Descripción de Recursos) es un lenguaje de sentencias afirmativas dirigido a expresar proposiciones utilizando precisos y formales vocabularios, en particular los especificados mediante RDFS [VOCABULARIO-RDF], se consideran como los bloques constitutivos de la Web Semántica (Yu, 2011). por ser un lenguaje para representar información sobre recursos en la World Wide Web los objetivos de diseño hacen énfasis en atributos como la generalidad y la precisión al expresar las proposiciones sobre algún tema, en lugar de la conformidad con cualquier modelo de procesamiento particular.

Figura 3-1: Ejemplo del bloque básico de construcción de una tripleta RDF



Fuente: Elaboración propia

3.1.4 SKOS

El Sistema simple de organización del conocimiento proporciona un modelo para expresar tanto el contenido de esquemas conceptuales como la estructura básica de listas de encabezamientos temáticos, tesauros, folksonomías, taxonomías, esquemas de

clasificación, y otros tipos similares de vocabulario controlado. Permite componer y publicar conceptos en la World Wide Web, enlazados con datos en la Web, Como aplicación del Resource Description Framework (RDF), SKOS e integrados en otros esquemas conceptuales

3.1.5 SKOS básico

En este esquema se identifican con URI los recursos conceptuales, están documentados con varios tipos de notas, semánticamente relacionados entre sí, etiquetados con cadenas en uno o más lenguajes naturales, agregados en esquemas conceptuales en jerarquías informales y redes de asociación.

3.1.6 SKOS avanzado

Para adaptarse a las necesidades de comunidades de práctica particulares o combinarse con otros vocabularios de modelado, el vocabulario de SKOS puede ampliarse a través de esquemas conceptuales y los recursos conceptuales pueden mapearse y agruparse en colecciones ordenadas o etiquetadas y las relaciones se pueden especificar entre etiquetas conceptuales.

3.1.7 OWL Lenguajes de Ontologías Web

El Lenguaje de Ontologías Web (OWL) facilita un mecanismo de interpretabilidad de contenido Web mejor que los mecanismos admitidos por XML, RDF, y esquema RDF (RDF-S). está diseñado para ser usado en aplicaciones que representan información para los humanos solamente y que en su lugar necesitan procesar el contenido de la información proporcionando vocabulario adicional junto con una semántica formal OWL. Este lenguaje cuenta con tres sublenguajes, con un nivel de expresividad incremental: OWL Lite, OWL DL, y OWL Full.

3.1.8 OWL Lite

Está diseñado para elaborar una clasificación jerárquica y restricciones simples. OWL Lite proporciona una ruta rápida de migración para tesauros y otras taxonomías. OWL Lite utiliza solo algunas de las características del lenguaje OWL y tiene más limitaciones en el

uso de las funciones que OWL DL o OWL Full. Por ejemplo, en OWL Lite las clases solo se pueden definir en términos de superclases (las superclases no pueden ser expresiones arbitrarias) y solo se pueden usar ciertos tipos de restricciones de clase. La equivalencia entre clases y relaciones de subclase entre clases también solo se permite entre clases con nombre. De forma similar, las restricciones en OWL Lite usan solo clases con nombre. OWL Lite también tiene una noción limitada de cardinalidad: las únicas cardinalidades que se permiten explícitamente son 0 o 1. OWL Lite tiene también una menor complejidad formal que OWL DL..(Bechhofer et al., 2004).

3.1.9 OWL DL

OWL DL está diseñado para obtener una máxima expresividad conservando completitud computacional (se garantiza que todas las conclusiones sean computables), y garantía de resolución en tiempos finito. OWL DL incluye todas las construcciones del lenguaje de OWL, pero sólo pueden ser usados bajo ciertas restricciones. OWL DL es denominado de esta forma debido a su correspondencia con la lógica de descripción (Description Logics, en inglés), un campo de investigación que estudia la lógica que compone la base formal de OWL.

3.1.10 OWL FULL

Tanto OWL DL como OWL Full usan el mismo vocabulario, aunque OWL DL está sujeto a algunas restricciones. Aproximadamente, OWL DL requiere separación de tipo (una clase no puede ser también un individuo o una propiedad, una propiedad tampoco puede ser un individuo o clase). Esto implica que las restricciones no se pueden aplicar a los elementos de lenguaje de OWL en sí (algo que está permitido en OWL completo). Además, OWL DL requiere que las propiedades sean `ObjectProperties` o `DatatypeProperties`: `DatatypeProperties` son relaciones entre instancias de clases y literales RDF y tipos de datos de esquema XML, mientras que `ObjectProperties` son relaciones entre instancias de dos clases. El documento Semántica y sintaxis abstracta de OWL explica las distinciones y limitaciones. Describimos el vocabulario completo OWL DL y OWL que amplía las construcciones de OWL Lite a continuación.

3.1.11 Lenguaje de Reglas de la Web Semántica (SWRL)

El Lenguaje de Reglas de la Web Semántica (SWRL, o Semantic Web Rule Language en inglés), se basa en una combinación de los sub-lenguajes OWL DL, OWL Lite y RuleML y usa una notación similar a la EBNF (Horrocks & Patel-Schneider, 2004). SWRL extiende el conjunto de axiomas de OWL para incluir cláusulas de Horn¹. Esto hace posible combinar cláusulas de Horn con una base de conocimientos en OWL. Además, posee una sintaxis abstracta de alto nivel, con la que se puede describir una ontología de OWL como una secuencia de axiomas y hechos.

Las reglas definidas con SWRL se definen como cláusulas de Horn, estableciendo una relación entre un antecedente y un consecuente, es decir, cuando se cumplen las condiciones definidas en el antecedente.

3.1.12 RIF Formato de intercambio de reglas

El Formato de intercambio de reglas fue encomendado por el World Wide Web Consortium en 2005 al Grupo de Trabajo respectivo para crear un estándar para el intercambio de reglas entre los sistemas de reglas, en particular entre los motores de reglas web. RIF se centró en el intercambio en lugar de tratar de desarrollar un único lenguaje de reglas único porque, a diferencia de otros estándares de la Web Semántica, como RDF, OWL y SPARQL, lo cual en ese momento permitió ver que un solo idioma no supliría las necesidades de muchos paradigmas populares para el uso de reglas en la representación del conocimiento y el modelado de negocios. Los sistemas de reglas conocidos se dividen en tres amplias categorías: primer orden, programación lógica y reglas de acción. Estos comparten poco en cuanto a la sintaxis y la semántica debido a sus grandes diferencias incluso entre sistemas de la misma categoría por lo que el intercambio de reglas ha sido reconocido como un objetivo de logros escasos.

3.1.13 SPARQL

Es un lenguaje de consulta estándar que permite la recuperación de RDF y RDFS. Esta tecnología permite que los usuarios obtengan la información que requieren sin considerar

la base de datos que se está utilizando o el formato para almacenar los datos. Esta herramienta se basa en la comparación de patrones gráficos los cuales están basados en tripletas como las usadas en RDF pero con la opción de que posee una variable consulta que reemplaza el sujeto, objeto y predicado en el término RDF (Aliprand & Unicode Consortium., 2003). Una consulta SPARQL se representa sintácticamente mediante un bloque que consiste en un formulario de consulta (SELECT, CONSTRUCT o DESCRIBE), cero o más cláusulas de conjunto de datos (FROM y FROM NAMED), una cláusula WHERE z, y posiblemente modificadores de solución (por ejemplo, DISTINCT). La cláusula WHERE proporciona un patrón de gráfico para hacer coincidir con el conjunto de datos RDF construido a partir de las cláusulas del conjunto de datos.

3.2 Conceptos

3.2.1 Datos Enlazados

La Web Semántica es una Web de Datos, de fechas y títulos y números de partes y propiedades químicas y cualquier otro dato que uno pueda concebir. La colección de tecnologías de la Web Semántica (RDF, OWL, SKOS, SPARQL, etc.) proporciona un entorno donde la aplicación puede consultar esos datos, hacer inferencias utilizando vocabularios, etc. Sin embargo, para hacer realidad la Web of Data, es importante tener la gran cantidad de datos en la Web disponibles en un formato estándar, accesible y manejable por las herramientas de la Web Semántica. Además, la Web Semántica no solo necesita acceder a los datos, sino que las relaciones entre los datos también deberían estar disponibles para crear una Web de Datos (en lugar de una colección pura de conjuntos de datos). Esta colección de conjuntos de datos interrelacionados en la Web también puede denominarse Datos Vinculados.(Neo4j Inc. et al., 2017)

Para lograr y crear datos vinculados, las tecnologías deben estar disponibles para un formato común (RDF), para hacer una conversión o sobre la marcha de acceso a las bases de datos existentes (relacionales, XML, HTML, etc.). También es importante poder configurar puntos finales de consulta para acceder a esos datos de manera más

conveniente. W3C proporciona una paleta de tecnologías (RDF, GRDDL, POWDER, RDFa, las próximas R2RML, RIF, SPARQL) para acceder a los datos.

3.2.2 Ontologías o Vocabularios

El W3C señala que gracias a que no existe una división clara entre lo que se denomina "vocabularios" u "ontologías" debería reservarse la denominación de ontología a un conjunto de términos complejo ya que los que se denominan "vocabularios" en la práctica pueden ir desde lo muy simple (describiendo solo uno o dos conceptos). a lo más complejo (con varios miles de términos)

Estos elementos que definen las relaciones y los conceptos (conocidos también como "términos") son usados para caracterizar relaciones posibles y definir restricciones posibles al uso de esos términos, por lo que en lo relacionado a las técnicas de inferencia son indispensables para la Web Semántica. (World Wide Web Consortium (W3C), 2017)

W3C ofrece una gran variedad de técnicas para describir y definir diferentes formas de vocabularios en un formato estándar. Ya que algunas aplicaciones pueden necesitar ontologías con complejos procedimientos de razonamiento que logren clasificar los términos para su uso en una aplicación en particular mientras que otros necesitarán un acuerdo sobre terminologías comunes, sin ningún rigor impuesto por un sistema lógico, algunas otras. Todo dependerá de los requisitos y los objetivos de las aplicaciones con lo cual además de incluir RDF y los esquemas RDF, el lenguaje de ontología web (OWL), el formato de intercambio de reglas (RIF) y sistema de organización de conocimiento simple (SKOS).

3.3 Componentes

3.3.1 Editor de Ontologías Protege

En sus inicios fue descrito como un sistema para soportar las recomendaciones de Lenguaje Web Ontológico OWL recomendado por el World Wide Web Consortium W3C (Musen, 2015)

y luego de incorporar los aportes del proyecto CO-ODE (Alan Rector et al., 2009) de la Universidad de Manchester. Actualmente se describe como un editor de ontología y base para construir sistemas inteligentes de código abierto y sin costo. Desarrollado en el Centro de Investigación en Informática Biomédica (BMIR) en la Facultad de Medicina de la Universidad de Stanford y cuenta con el respaldo de una fuerte comunidad de usuarios académicos, gubernamentales y corporativos, que usan Protege para desarrollar soluciones en áreas tan diversas como la biomedicina, el comercio electrónico y el modelado organizacional.(Holger Knublauch et al., 2017). La versión empleada en el desarrollo del presente trabajo es la 5.0.0.

3.3.2 Contenedores de RDF

La siguiente tabla relaciona los gestores de almacenamiento o contenedores de RDF

Tabla 3-3 Gestores de almacenamiento o contenedores de RDF.

NOMBRE	PLATAFORMA	TECNOLOGIAS	LENGUAJES DE CONSULTA
4store	Unix, Mac OS X	RDF	SPARQL
Bigdata	Windows, Mac, Linux y Solaris	RDF, RDFS, OWL Lite	SPARQL
Mulgara	Windows, UNIX, Linux, Solaris, Mac OS X e IRIX	RDF	ITQL, Futuro soporte para SPARQL
Virtuoso	Windows, Mac OS X, Linux y UNIX	RDF, XML	ISQL, SPARQL

3.3.3 Lenguajes de Consulta

Varios lenguajes han sido desarrollados para consultar información en grafos como Cypher (Neo4j Inc. et al., 2017) que opera sobre una Base de datos NoSQL propietaria que requiere APIS adicionales para interoperar con RDF y OWL que son los dos lenguajes

dominantes de la Web Semántica, por ende SPARQL es el lenguaje de consulta por defecto para RDF (Angles & Gutierrez, 2008).

3.3.4 Sistemas para extracción de información

- Procesamiento de lenguaje natural

El Procesamiento del Lenguaje Natural (PLN) es una de las bases de la inteligencia artificial IA y ha desempeñado múltiples papeles en este contexto, y su importancia dentro de este campo ha crecido y decrecido a consecuencia de cambios tecnológicos y científicos (Jaime Carbonell, 1992). Los primeros intentos de traducir textos por ordenador a finales de los cuarenta y durante los cincuenta, fracasaron debido a la escasa potencia de los ordenadores y a la escasa sofisticación lingüística. Sin embargo, los esfuerzos realizados en décadas más recientes han permitido resurgir la investigación en esta área que ha conducido a sistemas susceptibles de ser explotados industrialmente. Estos progresos favorables se deben a una combinación de factores que van desde un enorme aumento en la potencia de los ordenadores en relación a su coste hasta modelos del lenguaje humano mejores y más susceptibles de ser tratados computacionalmente. Por otra parte, cada vez se evidencia más la necesidad de sistemas de PLN para procesar datos textuales, incluyendo traducción, clasificación, recuperación y extracción de información.

- Sistema de recomendación semántico adaptable al contexto

Los sistemas de recomendación semánticos basan su proceso de recomendación sobre una base de conocimiento, normalmente definida a través de un esquema de conceptos (como una taxonomía o un tesoro) o una ontología, y además, para ser considerados como adaptables al contexto, estos tipos de sistemas deben tomar en consideración diferentes factores (temporales, de lugar, nivel de experiencia del usuario, dispositivo que se está utilizando en el momento de recibir la recomendación, entre otros.) para inferir el contexto en que se encuentra el usuario y adaptar las recomendación a esas circunstancias, facilitando el acceso de los usuarios a la información que necesitan (E. Peis; J. M. Morales-del-Castillo, 2008).

Para realizar este proceso de inferencia, las técnicas utilizadas por los sistemas de recomendación adaptables al contexto usan lo que se considera como conocimiento funcional, es decir, se tiene conocimiento acerca de cómo un elemento en particular responde a una necesidad particular del usuario, y por lo tanto se puede razonar sobre la relación entre una necesidad y una posible recomendación (Burke, 2002).

- Ontología de contexto multidimensional

El contexto es cualquier información que puede ser usada para caracterizar la situación de una entidad, donde una entidad puede ser una persona, lugar u objeto físico o computacional, por lo que puede verse como un espacio multidimensional donde cada dimensión es representada por una ontología específica (González Franco & Alejandres Sánchez, 2015). Para ello, se maneja una red de ontologías, relacionadas entre sí, las cuales en conjunto contienen la información que describe el contexto y adquieren el nombre de ontología de contexto multidimensional.

3.3.5 End-point SPARQL, Punto final de acceso

Un punto final SPARQL es un servicio de protocolo SPARQL que permite a los usuarios (no solo humanos) consultar una base de conocimiento a través del lenguaje SPARQL. Los resultados se devuelven típicamente en uno o más formatos procesables por máquina. Por lo tanto, un punto final SPARQL se concibe principalmente como una interfaz amigable para la máquina hacia una base de conocimiento. Para el presente trabajo se ha seleccionado el framework de Web Semántica de código abierto Apache Jena, el cual está escrito en Java y proporciona una API para extraer datos y escribir en gráficos RDF. Los gráficos se representan como un "modelo" abstracto. Un modelo se puede obtener con datos de archivos, bases de datos, URL o una combinación de estos. Este framework cuenta con un módulo que permite disponer las tripletas en un end-point SPARQL accesible a través de HTTP denominado Fuseki que proporciona una interacción al estilo REST para los datos RDF.

3.4 Clasificaciones de sistemas de recomendación

Existen dos clasificaciones de sistemas de recomendación manejadas en la literatura, la primera de ellas define las siguientes categorías:

3.4.1 Basados en contenido

En esta clase de sistemas al usuario le son recomendados ítems basados en lo que le ha gustado previamente. En (Maynard, Roberts, Greenwood, & Bontcheva, 2015) se presenta una plataforma para almacenar y analizar contenido de redes sociales de gran volumen. El marco analítico en tiempo real comprende la anotación semántica, los datos abiertos vinculados, la búsqueda semántica y los componentes de agregación de resultados dinámicos. Con interfaces de búsqueda semántica y visualización de la información, como nubes de términos, matrices de coincidencia y mapas arbóreos, los usuarios pueden analizar y conservar los resultados. A continuación, se revisan algunos trabajos realizados en esta clase de sistemas.

Una plataforma llevada a la práctica a través de dos casos de estudio donde por un lado se investiga la actitud hacia el cambio climático público, a través de su involucramiento con los temas ambientales; y otro caso donde se analizan los trinos de los políticos en el Reino Unido a través de esta red social y la respuesta correspondiente del público durante las elecciones de 2015 en el Reino Unido. También se presenta una breve evaluación y discusión de algunos de los componentes principales del análisis de texto, específicamente ajustados logrando demostrar la escalabilidad y la eficiencia en la aplicación para los casos estudiados.

Otro trabajo de esta clase presenta la solución a un problema puntual en el almacenamiento e indexación de documentos producidos por procesos realizados en una entidad de carácter público en el sector de la educación (Abello, 2015), debido a que la información posible de extraer de los mismos es limitada. La solución presentada gira en torno a la obtención de una base de conocimiento, atendiendo los lineamientos y la

estructura de datos enlazados, que determina la relevancia para su publicación para ser consultada y tenida en cuenta en la entidad para la toma de decisiones.

El método propuesto para la extracción de información partiendo de documentos semiestructurados será el sustento para la generación de la base de conocimiento, y punto de partida para desplegar el prototipo que a través de cuatro fases extrae la información necesaria incluyendo la fase la inferenciación a través de la Persistencia de conocimiento.

3.4.2 Colaborativos

Esta clasificación agrupa a los sistemas de recomendación en basados en reglas, de filtrado colaborativo y de personalización basada en contenidos de este modo al usuario se le recomiendan ítems que han sido del agrado de usuarios con preferencias similares. Como muestra de este tipo de sistemas se cuentan sistemas que proporcionan una alternativa de mejora para la integración de los recursos científicos, a partir de conceptos y tecnologías de la Web Semántica, con énfasis en conceptos de datos enlazados. En (Quiroga, 2015) se presenta un prototipo mediante el uso de sets de datos en el ámbito de la producción académica e investigativa junto con la definición y desarrollo de ciclo de vida propuesto para la producción de datos enlazados enmarcado en una revisión amplia de tecnologías y herramientas para su ejecución.

3.4.3 Híbridos

Estos sistemas combinan el enfoque basado en contenidos con el enfoque colaborativo, en (DAU & Andrews, 2013) se busca complementar los mejores elementos del Inteligencia de Negocios convencional con las tecnologías semánticas más recientes de la Web Semántica en forma de RDF. El objetivo de la Plataforma Europea Framework Siete denominado CUBIST es proporcionar a los usuarios finales "análisis visuales conceptualmente relevantes y fáciles de usar" para permitirles explorar sus datos de nuevas maneras, descubriendo significados ocultos y resolviendo problemas complejos. En este empeño tres de las temáticas en CUBIST fueron casos de uso como la industria espacial, biología computacional y consultoría de reclutamiento. Cada caso de uso proporcionó requisitos específicos y evaluó sus resultados.

4 Desarrollo de un Sistema de recomendaciones para el fortalecimiento de competencias laborales a profesionales recién egresados usando Ontologías y RDFs

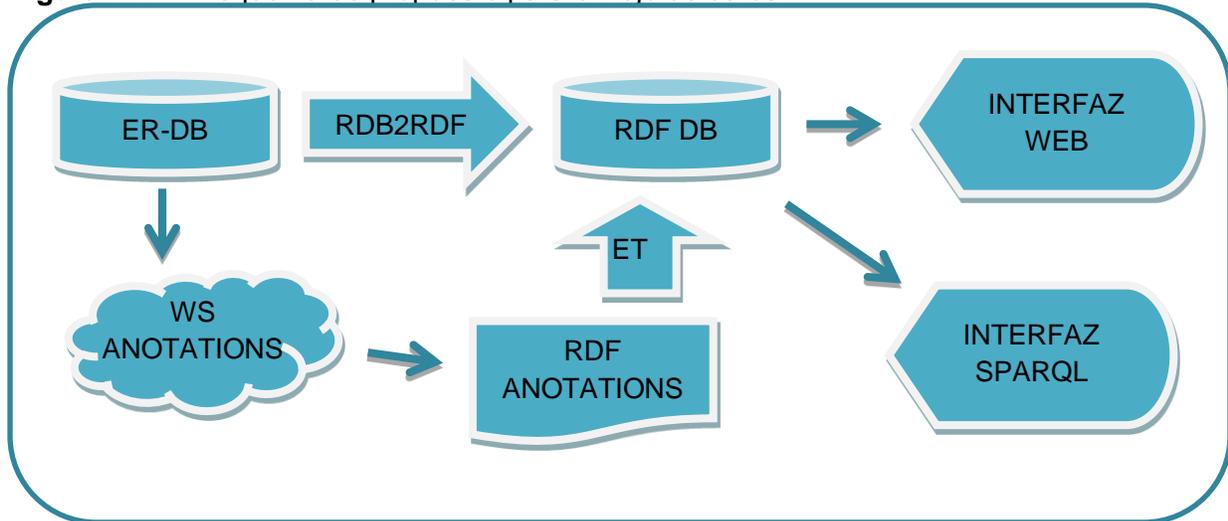
El contenido de este capítulo documenta el desarrollo sistema para generar información que pueda ser útil para egresados que hacen uso de los servicios de la bolsa de empleo de la Universidad Nacional obteniendo recomendaciones basadas en información resultante de consultas sobre tripletas en formato RDF con base en un esquema semántico obtenido del procesamiento de lenguaje natural del perfil de las hojas de vida seleccionadas por empresas, a continuación, se detalla el esquema que guiará el desarrollo del presente trabajo.

4.1 Esquema de propuesto para el flujo de datos

El esquema general propuesto inicia con el pre-procesamiento de datos que permiten extraer la información de la base de datos para convertirla a tripletas y ser almacenadas en un repositorio de base de datos compatible con la tecnología semántica necesaria. Con el fin de enriquecer el modelo RDF del repositorio se hace uso de herramientas para procesamiento en lenguaje natural aplicada sobre una variable de tipo texto en la que los egresados detallan su perfil y brindan información no estructurada dirigida a los empleadores en que expresan básicamente sus logros su experiencia y su trayectoria académica o laboral, dicha información de egresados cuyos perfiles han sido seleccionadas por parte de las empresas, será procesada usando técnicas de extracción

en lenguaje natural para la obtención modelos individuales por cada perfil en formato RDF el cual a través un procesamiento con herramientas ETL son cargadas e integrados al repositorio de base de datos central.

Figura 4-1: Esquema de propuesto para el flujo de datos.



Fuente: Elaboración propia

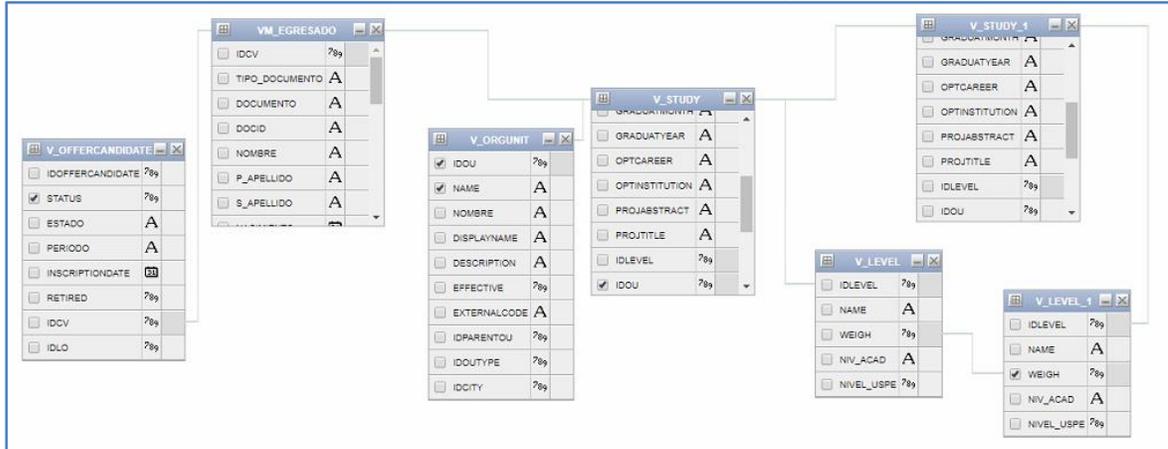
Al contar con los datos centralizados en formato RDF se despliegan dos interfaces que permiten ejecutar consultas en formato Sparql y brindar a los oferentes y demandantes de vacantes información adicional que en un sentido ayude a orientar a las empresas a identificar perfiles adecuados con base en ciertas habilidades que en ocasiones no están almacenadas en los campos de la base de datos relacional destinados para tal fin y en otro sentido orientar a los egresados a identificar posibles estudios complementarios para el enriquecimiento de su currículum.

4.2 Conversión de datos relacionales a tripletas RDB2RDF

Actualmente el Sistema de Información almacena los datos en un esquema de base de datos Base de Datos Oracle licenciado en la versión 11gR2 con lo cual y teniendo en cuenta las prestaciones que dicha herramienta presta a la tecnología Web semántica se realizó un análisis en el que se generaron vistas como producto de una consulta (Lei Chen & Nansheng Yao, 2010) donde se priorizaron variables significativas bajo la óptica de

tripletas de Sujeto Objeto y Predicado, con lo cual además se reduce el número de llaves primarias internas que carecen de valor analítico.

Figura 4-2: Consulta realizada con utilidad Query designer de Oracle APEX, sobre el modelo ER de la BD transaccional que define la vista usada para el proceso RDB2RDF.



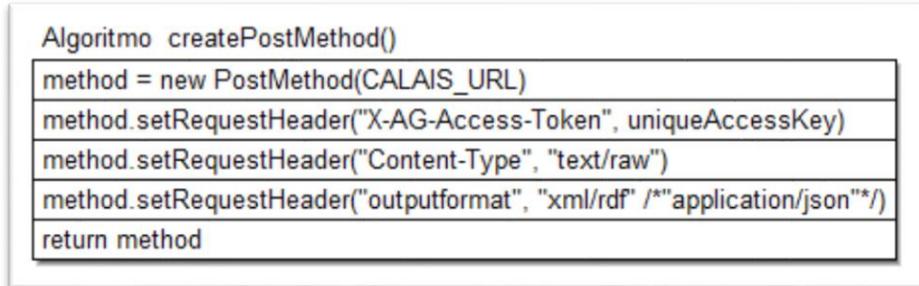
Fuente: Elaboración propia, adaptado de base de datos del Sistema de Información de Egresados.

4.3 Texto a RDF

Con el fin de convertir la información de texto contenida en el campo de base de datos destinada para tal fin y teniendo en cuenta que los objetos almacenados en la base de datos destinados para tal fin, tanto en las versiones 11gR2 o 12cR2 dependen de un extractor externo como el proporcionado por GATE que es la sigla en inglés de la denominada Arquitectura General para Ingeniería en Textos que es un procesador de lenguaje natural de código abierto y extractor de información(Cunningham, 2017) o el Open Calais proporcionado por Thomson Reuters que mediante llamadas a Servicios Web permite analizar hasta 5000 documentos diarios obteniendo los RDFs en formato XML (Loria, Gardent, & Loria, 2015), junto con la ventaja de incorporar flexibilidad en el manejo de varios idiomas entre ellos el español. Las ventajas proporcionadas por la interfaz de programación de aplicaciones API serán usadas en esta etapa. A continuación, se detallan los métodos principales involucrados en el proceso de extracción provistos por la herramienta seleccionada y se visualizan sus algoritmos mediante estructogramas. Los respectivos códigos se encuentran en el anexo 2:

- Implementación del método Post y especificación de los parámetros de trabajo, de los cuales se destaca la selección del formato que para el presente caso será xml/rdf, la url del servicio y los datos de autenticación.

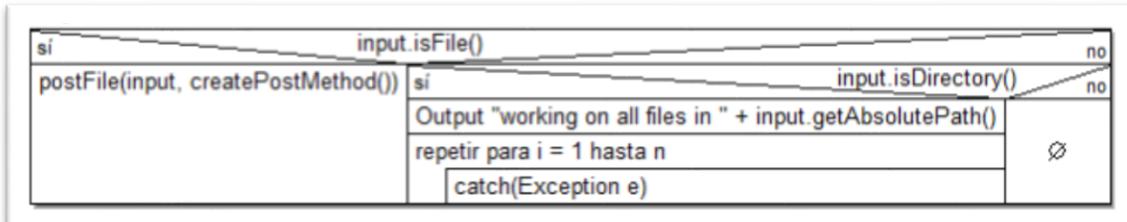
Figura 4-3: Estructograma del método de parametrización vía Post createPostMethod ()



Fuente: Elaboración propia, adaptado de Open Calais (anexo 2).

- Método run () que permite seleccionar un archivo o más a través del directorio que los contiene, para el presente caso de usa un directorio donde se ubican los archivos generados por un flujo ETL desarrollado para tal fin y que se presentará en próximos numerales.

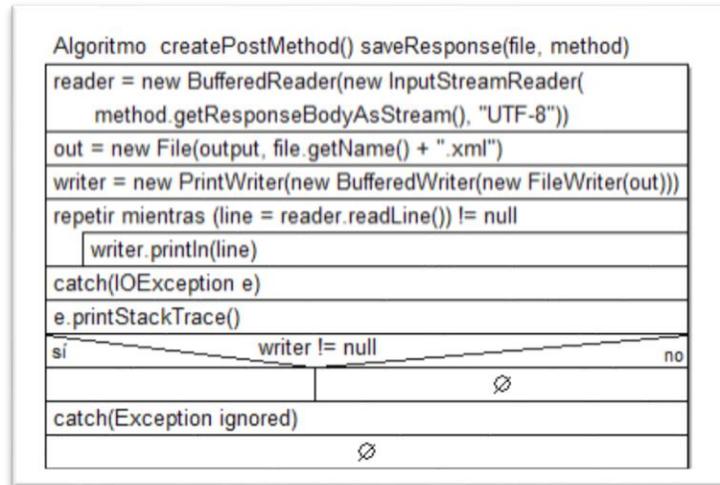
Figura 4-4: Estructograma del método de selección de archivos run ()



Fuente: Elaboración propia, adaptado de Open Calais (anexo 2).

- Método saveResponse (File file, PostMethod method) que almacena las respuestas obtenidas del Servicio Web.

Figura 4-5: Estructograma del método saveResponse (File file, PostMethod method)



Fuente: Elaboración propia, adaptado de Open Calais (anexo 2).

Teniendo en cuenta la documentación del RESTful API para el extractor seleccionado y la parametrización requerida, se observa para la clase cliente (ver Figura 4-2) que sus métodos requieren un directorio con los ficheros en formato de texto para cada uno de los perfiles.

Figura 4-6: Modelo UML de la clase cliente que invoca el Servicio Web de etiquetado automático.

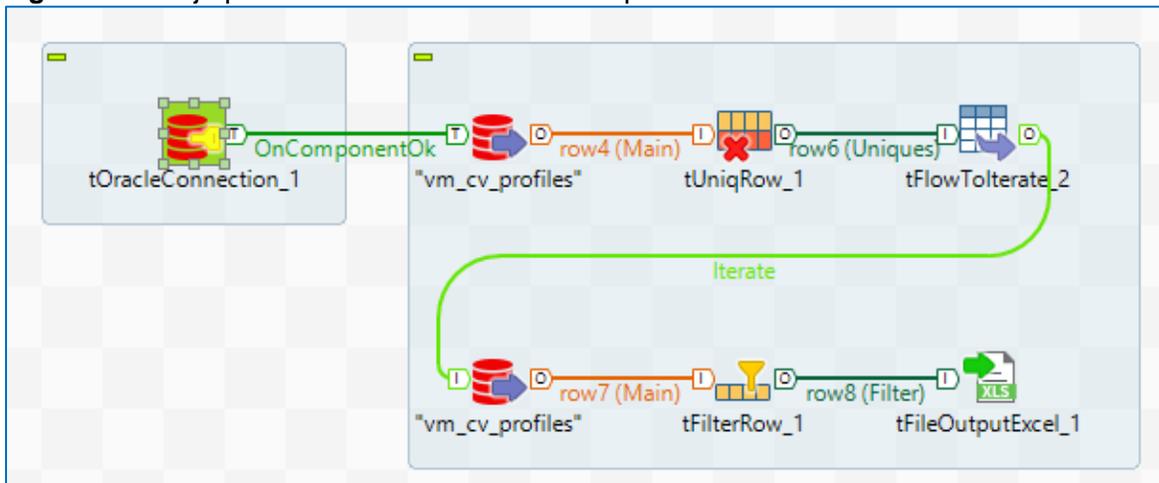


Fuente: Elaboración propia, adaptado de Open Calais (anexo 2).

4.3.1 Extracción de datos de perfiles seleccionados por empresas oferentes

Teniendo en cuenta los requisitos para la implementación se generó un flujo para la extracción automática en el formato indicado Mediante el ETL Talend Open Studio para integración de Datos con el que se realizaron las iteraciones correspondientes sobre la tabla que almacena la información de perfil requerido ver Figura 4-6:

Figura 4-7: Flujo para la extracción de datos de perfiles a archivos en formato .xls



Fuente: Elaboración propia

A continuación, se detalla la funcionalidad de cada uno de los componentes que intervienen en el proceso de extracción.

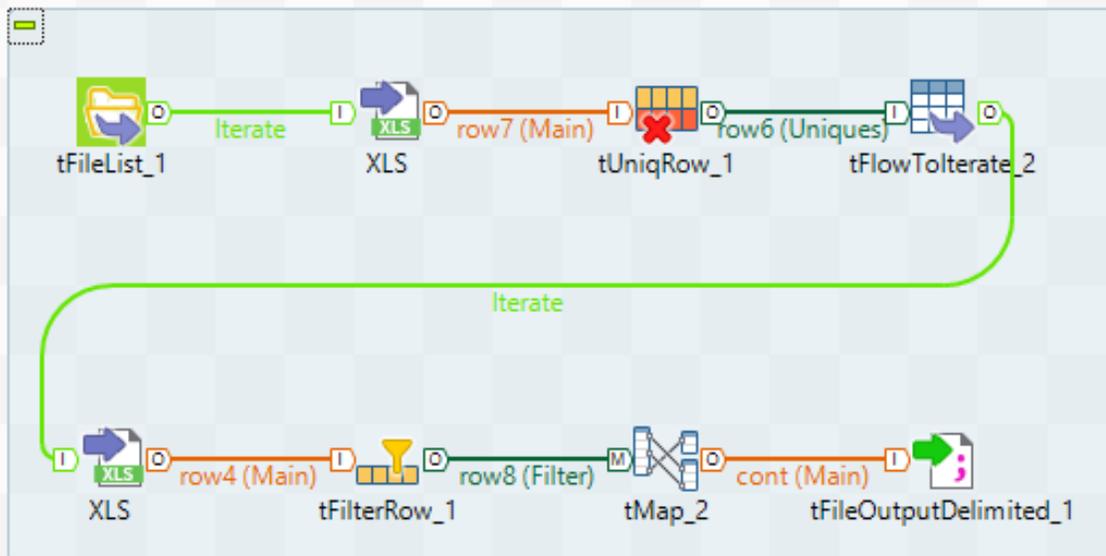
Tabla 4-1 Descripción detallada del flujo de datos para la preparación de los datos de perfiles.

Componente	Funcionalidad
tOracleConnection_1	Conecta con la base de datos de producción del sistema transaccional, dicha conexión se realiza a través de un Identificador de sistema SID por puerto 1521.
vm_cv_profiles	Consulta los perfiles junto con una llave que permitirá la identificación de los resultados.
tUniqRow_1	Permite que el registro leído hasta ese punto se convierta en parámetro del siguiente flujo

tFlowTolterate_2	Desencadena una iteración de los procesos siguientes por cada registro recibido.
vm_cv_profiles	Consulta uno a uno los perfiles para ser almacenados.
tFilterRow_1	Limita el número de registros para el correcto manejo de idioma, definido por la herramienta.
tFileOutputExcel_1	Almacena los resultados en el formato .xls requerido.

Luego de procesar los datos extraídos, los registros resultantes son separados en múltiples archivos a un archivo con extensión .txt cuyo nombre corresponde identificador se asigna del mismo y perfil como contenido.

Figura 4-8: Flujo para la obtención de un único archivo de texto para cada uno de los perfiles.



Fuente: Elaboración propia

A continuación, se detalla la funcionalidad de cada uno de los componentes que intervienen en el proceso de extracción que en esta etapa permite la adecuación en al formato de archivos de texto requeridos.

Tabla 4-2 Descripción detallada del flujo de datos para el preprocesamiento de los datos de perfiles.

Componente	Funcionalidad
tFileList_1	Genera un listado de los archivos .xls almacenados en el directorio especificado y desencadena una iteración para cada uno de los archivos.
XLS	Lee el contenido de cada uno de los archivos en formato XLS.
tUniqRow_1	Permite que el registro leído hasta ese punto se convierta en parámetro del siguiente flujo
tFlowTolterate_2	Desencadena una iteración por cada registro recibido
XLS	Lee registro por registro el contenido de cada uno de los archivos en formato XLS.
tFilterRow_1	Limita a 1 el número de registros para el correcto procesamiento de la herramienta.
tMap_2	Excluye el identificador dentro contenido del archivo de texto
tFileOutputDeliited_1	Almacena los archivos con el identificador como nombre y perfil como contenido.

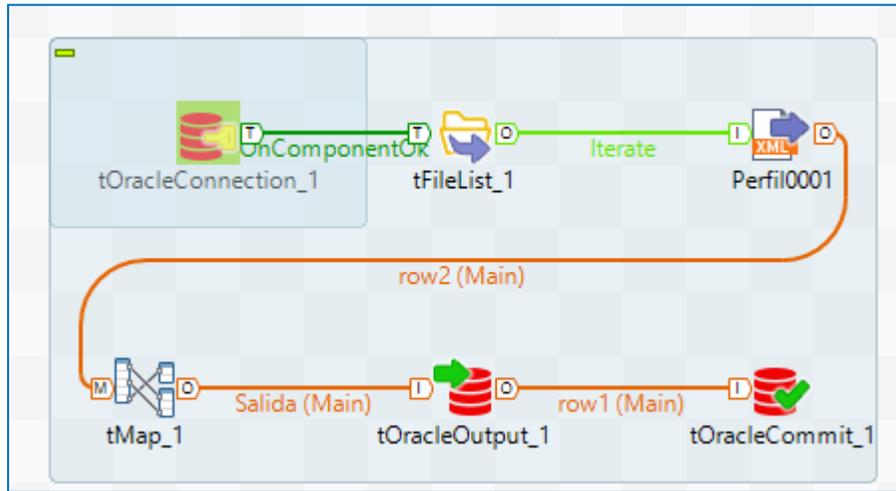
En este punto ya contamos con un archivo de texto por cada uno de los perfiles que serán procesados por la herramienta seleccionada.

4.3.2 Obtención de RDFs de perfiles y almacenamiento de etiquetas

Al contar con los archivos en el formato necesario se efectúa la ejecución del código indicado para lo cual se compila previamente con el JDK 1.7.

Luego de ejecutar el código proporcionado y de obtener los archivos XML generados se llevó a cabo un flujo con el que son identificadas las tripletas correspondientes a las anotaciones o TAGs correspondientes, y almacenadas en la tabla RDF asignada.

Figura 4-9: Flujo para la obtención y almacenamiento de las etiquetas obtenidas en el RDF.



Fuente: Elaboración propia

A continuación, se detalla la funcionalidad de cada uno de los componentes que intervienen en el proceso de carga de los datos RDF.

Tabla 4-3 Descripción detallada del flujo de datos para el almacenamiento de etiquetas.

Componente	Funcionalidad
tOracleConnection_1	Conecta con la base de datos con funcionalidades RDF habilitadas, dicha conexión se realiza a través de un Identificador de sistema SID por puerto 1521.
tFileList_1	Genera un listado de los archivos contenidos en el directorio especificado.
Perfil0001	Extrae información de etiquetas específicas para cada uno de los perfiles.
tMap_1	Especifica los campos etiquetados que serán incluidos en la salida.
tOracleOutput_1	Inserta la información en la tabla correspondiente.
tOracleCommit_1	Confirma las inserciones sobre la Base de Datos.

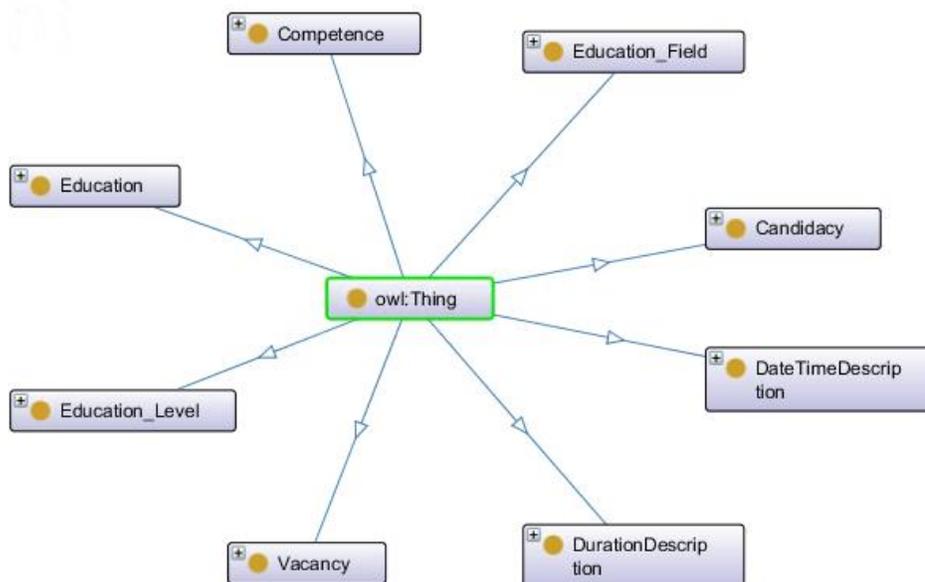
Basado en los resultados y con el fin de procurar una mejora en los mismos se repitió el procedimiento incorporando un paso previo adicional para realizar el etiquetado en el

idioma inglés haciendo uso de las prestaciones de la página de traducción en línea www.freetranslation.com que permite traducción de formatos en Excel logrando una mejora en la cantidad de tripletas obtenidas.

4.4 Modelado de datos RDF

Para el almacenamiento se debe seguir una jerarquía de lo físico a lo lógico en cuanto a tablas y modelos, para el caso particular el desarrollo de la ontología se modeló con base en la ontología de código abierto EducationOntology.owl disponible en el repositorio del trabajo Gestión de Ontologías para Recursos Humanos (Boris Villazón-Terrazas, 2014) ver imagen. Se seleccionó dicha ontología ya que hace parte de un conjunto de ontologías interrelacionadas con otras ontologías del ámbito laboral como son competencias, actividades económicas, ofertas de trabajo, habilidades de trabajo entre otras que dan posibilidades de ampliar el presente trabajo a futuros desarrollos que amplíen las capacidades del presente trabajo.

Figura 4-10: Representación gráfica de la ontología adaptada.



Fuente: Elaboración propia, adaptado de EducationOntology.owl

Una vez almacenados las tripletas en la tabla asociada al modelo predefinido dichos datos son procesados mediante la función incorporada en el motor de la Base de datos

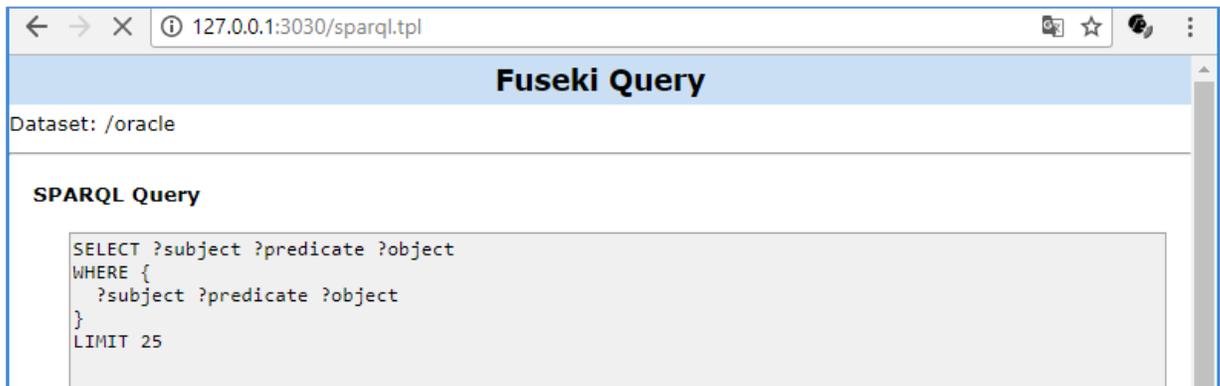
sem_rulebases. Dicha función cuenta con un parámetro que permite elegir entre inferencias obtenidas mediante RDFS o OWL2RL, y que en cualquier caso las nuevas tripletas inferidas se integrarán al repositorio.

4.5 Interfaces

4.5.1 SPARQL EndPoint

Al contar con las tripletas almacenadas en la base de datos, se implementó un servidor de consultas SPARQL basado en Apache Jena Fuseki 2.4 que proporciona funcionalidades de actualización mediante REST, consulta y actualización de SPARQL utilizando el protocolo SPARQL sobre HTTP.

Figura 4-11: Vista del SPARQL EndPoint del servidor SPARQL Apache Jena Fuseki 2.4 implementado.



Fuente: Elaboración propia

Dicha herramienta se implementa en entorno de desarrollo para validaciones y pruebas necesarias en la etapa de diseño de las interfaces de consultas para usuarios finales.

Figura 4-12: Salida de una consulta de prueba del SPARQL EndPoint implementado.

```

subject
predicate                                     | object
-----
<http://www.unal.edu.co/bolsaempleo/EstSecundarios/Ingenieria-De-Sistemas>
<http://www.unal.edu.co/bolsaempleo/EstSecundarios> | <Especialista-En-Seguridad-De-La-Informacion>
<http://www.unal.edu.co/bolsaempleo/EstSecundarios/Ingenieria-Industrial>
<http://www.unal.edu.co/bolsaempleo/EstSecundarios> | <Especialista-En-Gerencia-Logistica>
<http://www.unal.edu.co/bolsaempleo/EstSecundarios/Farmacia>
<http://www.unal.edu.co/bolsaempleo/EstSecundarios> | <Especialista-En-Gerencia-Logistica>

```

Fuente: Elaboración propia

4.5.2 Formulario de consultas

Para la elaboración de los formularios de consulta habituales se llevaron a cabo los análisis del caso de uso siguiente que de modo genérico permiten a los gestores de la bolsa de empleo contar con las herramientas para orientar tanto a empresas como a egresados se implementa en entorno de desarrollo para validaciones y pruebas necesarias en la etapa de diseño de las interfaces de consultas para usuarios finales.

Figura 4-13: Diagrama de caso de uso para la interacción con la interfaz de visualización de recomendaciones.

Fuente: Elaboración propia

Las interfaces fueron generadas con la herramienta de desarrollo Oracle Application Express o APEX es una herramienta RAD desarrollo rápido de aplicaciones, que se ejecuta con una base de datos Oracle, y que junto con la función SEM_MATCH integrada en el DBMS (Sistema de Administración de Bases de Datos) permite a la herramienta manejar

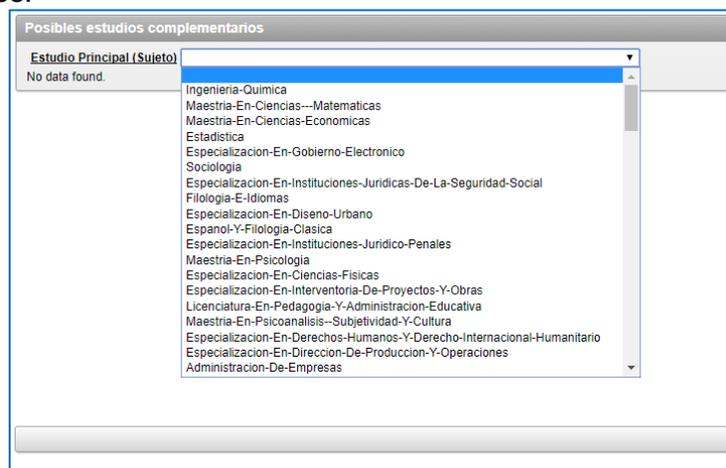
resultados de consultas SPARQL como tablas integradas a sentencias de consultas SQL(Oldag, 2014).

4.6 Resultados de las pruebas

Las interfaces logradas corresponden a los casos de uso principales que corresponden inicialmente a la consulta de estudios complementarios basados en los estudios de currículos que han pasado a etapas de selección por las empresas en las convocatorias de la bolsa de empleo.

La interfaz inicialmente permite la selección de estudios denominados principales que corresponde básicamente al título más reciente que les permite ser registrados en la base de datos relacional y que al ser transformados en tripletas constituyen el sujeto cuyo predicado indica que cuenta con estudios complementarios que serán desplegados como resultado de dicha selección inicial.

Figura 4-14: Interfaz para la selección de programas académicos para obtención de recomendaciones.



Fuente: Elaboración propia

La segunda interfaz permite al gestor realizar la búsqueda de perfiles relacionados con temas específicos relacionados con habilidades puntuales específicas, obteniendo como resultado la relación de usuarios que cuentan con dicha habilidad o competencia específica.

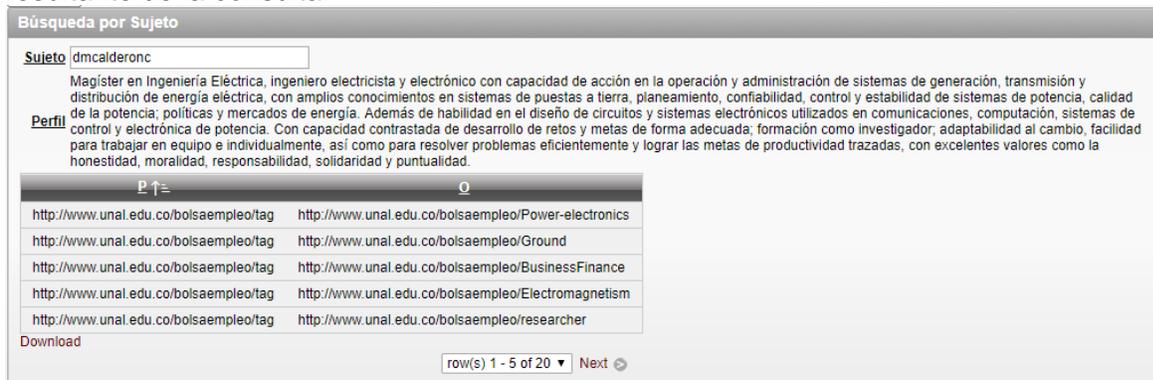
Figura 4-15: Interfaz para la selección de temáticas específicas para la obtención de perfiles opcionados.



Fuente: Elaboración propia

Una vez son desplegados los resultados puede ampliarse búsqueda a una vista as detallada del perfil que fue sugerido, junto con las demás tripletas obtenidas del proceso de etiquetado automático.

Figura 4-16: Interfaz que le permite al gestor de ofertas autorizado verificar el perfil resultante de la consulta.



Fuente: Elaboración propia

4.6.1 Evaluación experimental

Para la evaluación experimental se tomaron en consideración las temáticas presentes en ofertas laborales recientes, que a su vez son dadas como objeto de búsqueda en las tripletas correspondientes con las anotaciones o etiquetados de los perfiles; en cuanto a las recomendaciones de estudios complementarios se evaluó la pertinencia de las recomendaciones dadas para estudios teniendo en cuenta poder obtener estudios de igual o superior nivel al consultado, es decir que se verificó que la herramienta no recomiende un pregrado a alguien que ya cuenta con un título de posgrado.

Adicionalmente se describieron en la siguiente tabla algunos de los atributos de servicio que actualmente no están presentes en la bolsa de empleo y que se observa que podrían complementarse para mejorar las cifras de empleos obtenidos por los egresados de la Universidad.

Tabla 4-4 Actividades de servicio y otras especificaciones operativas.

Actividades de servicio y otras especificaciones operativas (Alter, 2012)					
Actividad	Cientes participantes	No-Cientes participantes	Entidades de Información	Entidades Tecnológicas	Producto/ Servicio
Selección de Estudios Principales	Egresado	Gestor de Bolsa de Empleo	Repositorio de tripletas de estudios complementarios recomendados	Interfaz de consulta de recomendaciones de estudios complementarios	Recomendación de estudios complementarios
Selección de habilidades / competencias	Empresas	Gestor de Bolsa de Empleo	Repositorio de tripletas de anotaciones	Interfaz de consulta de recomendaciones de perfiles afines	Perfiles recomendados

A continuación, se presentan análisis de los resultados a casos específicos propios del funcionamiento habitual de la bolsa de empleo:

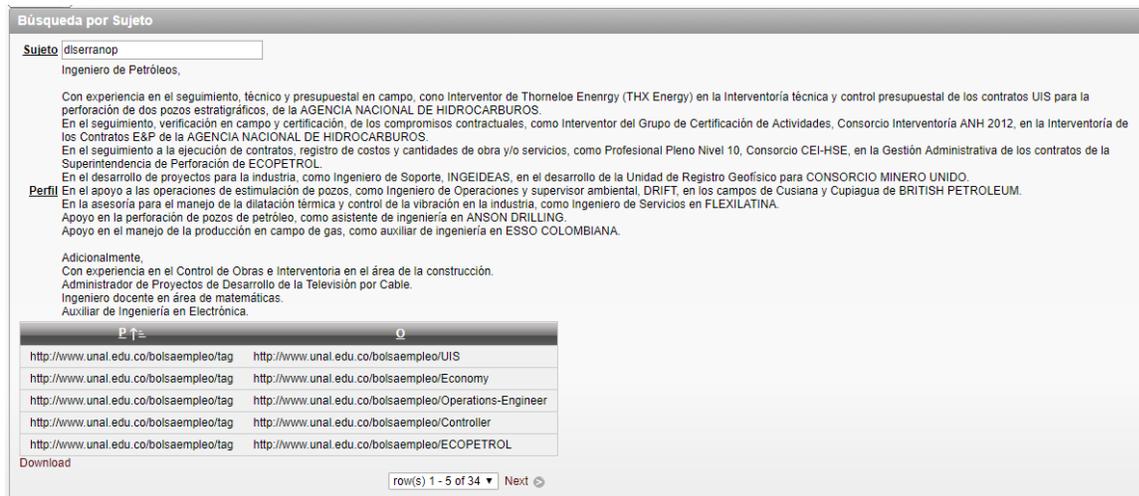
Caso 1: Requerimiento por parte de una empresa de un Ingeniero de Soporte:

Resultado: se obtienen 5 perfiles de egresados que cumplen con la condición inicial.

Figura 4-17: Resultados de caso de prueba para atender solicitud de empresa.

Fuente: Elaboración propia

Con los resultados obtenidos el sistema permite verificar la información detallada de cada uno de los perfiles seleccionando alguno de los sujetos como se observa en la siguiente imagen.

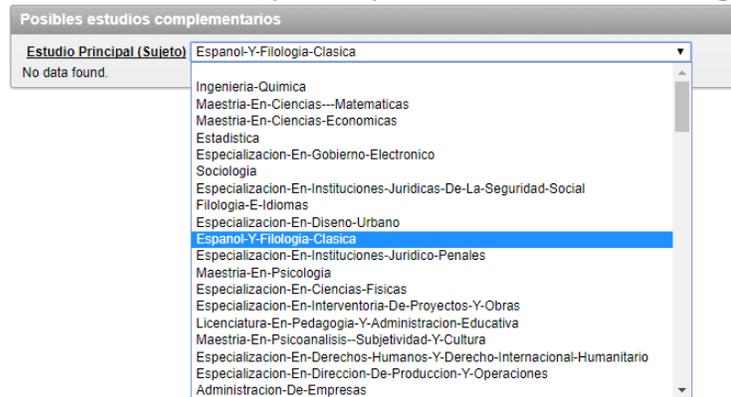
Figura 4-18: Resultados de caso de prueba para atender solicitud de empresa – Vista detallada.

Fuente: Elaboración propia

Caso 2: Requerimiento por parte de un Egresado de orientación de sobre estudios complementarios que pueda relacionar o realizar para actualizar su hoja de vida:

Resultado: Se obtienen la relación de estudios complementarios que cumplen con la condición inicial.

Figura 4-19: Resultados de caso de prueba para atender solicitud de egresado.



Fuente: Elaboración propia.

Los resultados muestran que la herramienta cumple con proveer la información solicitada, lo que permitirá contar como una alternativa adicional de atención a los usuarios.

5 Conclusiones y recomendaciones

5.1 Conclusiones

El campo de perfil destinado para realizar la descripción resumida de la trayectoria profesional del egresado incluye aptitudes o competencias específicas que no siempre se encuentran relacionadas en tablas destinadas para el registro de estudios o experiencia laboral, ya que los modelos entidad relación son muy restrictivos en cumplir con unos mínimos de duración, tiempo de desempeño, cargo, tipo de vinculación, Institución educativa etc. Por lo anterior se encuentra conveniente procesar dichos perfiles con tecnologías semánticas ya que permite consultar los datos en formatos RDF generado un punto en común complementario, que puede propiciar la coincidencia entre empresas y egresados dentro de los procesos de selección y contratación de la bolsa de empleo del caso abordado en el presente trabajo.

A pesar de que muchas de las tendencias actuales se orientan hacia tecnologías en la nube y el Big Data que requieren una transformación o migración más profunda a nivel de ciertas infraestructuras de las entidades, la Web semántica se mantiene como una solución integradora con una accesibilidad conveniente a los datos en un entorno como el colombiano donde se promueve la transparencia y la apertura de datos.

La revisión y evaluación que desde la academia se realiza a las herramientas que permiten habilitar las funcionalidades de la web semántica, muestran a lo largo del tiempo que se avanza en la consolidación y madurez de herramientas que integran elementos que permiten reducir el nivel de complejidad, haciendo que esta tecnología sea más fácil de

implementar principalmente en el sector público donde la austeridad institucional no excluye las áreas de TI.

Teniendo en cuenta la alta disponibilidad de recursos ontológicos se hace conveniente la reutilización de estos metadatos ya que en la medida que estos sean adaptados en lugar de redundar en nuevos desarrollos, se puede facilitar una mayor interoperabilidad desde la óptica de los datos enlazados.

5.2 Recomendaciones

Se recomienda que la periodicidad para realizar la consolidación y procesamiento de datos se dé inicialmente de modo semestral complementado con un análisis anual que permita comparar las recomendaciones dadas frente a los nuevos estudios registrados por los egresados en dicho lapso de tiempo.

Como trabajo posterior se puede pensar en evaluar desde el punto de vista de procesos como la incorporación de nuevos servicios de orientación para la remisión de currículos puede mejorar las tasas de empleabilidad de los egresados de la Universidad Nacional de Colombia.

Teniendo en cuenta las diferencias en los resultados obtenidos en el subproceso de etiquetado automático se puede abordar como trabajo futuro temas relacionados con el etiquetado automático en idioma español y su efectividad frente al idioma inglés.

A. Anexo 1: Servicios autorizados para la Bolsa de Empleo de la Universidad Nacional de Colombia.

RESOLUCIÓN NÚMERO 000605 DE 2017 HOJA N° 2

"Por medio de la cual se renueva a una Bolsa de Empleo de Gestión y Colocación de Empleo"

Que el Artículo 7 de la Resolución 3999 de 2015, indicó que para obtener autorización o renovación para la prestación de los servicios de gestión y colocación del Servicio Público de Empleo, se deberá planificar técnica y operativamente la prestación de los servicios, los recursos físicos, financieros y tecnológicos, todos estos sustentados a través de un proyecto de viabilidad.

Que de acuerdo con el párrafo 3° del Artículo 2.2.6.1.2.19 del decreto 1072 de 2015, los prestadores autorizados deberán mantener durante el tiempo que se haya concedido la autorización, las condiciones jurídicas, operativas y técnicas que le permitieron obtenerla. En caso de incumplimiento, esta Unidad podrá suspenderla o revocarla mediante acto motivado.

Que mediante la Resolución 293 del Ministerio del Trabajo, se modificó la caracterización de los puntos de atención para la prestación de los servicios de gestión y colocación de empleo.

Que mediante radicación No. SPE-GMSE-2017-ER-000514 de 28 de marzo de 2017, la doctora CLARA INÉS SABOGAL, actuando en calidad de Coordinadora Nacional del Programa de Egresados, de la **UNIVERSIDAD NACIONAL DE COLOMBIA** institución de carácter pública, identificada con NIT 899999063-3 y domicilio en Bogotá, solicitó renovación de autorización para la prestación del Servicio Público de Empleo como Bolsa de Empleo.

Que la **UNIVERSIDAD NACIONAL DE COLOMBIA**, dando cumplimiento a la normatividad vigente, presentó la siguiente documentación: a) Solicitud de renovación para la prestación de los servicios de Gestión y Colocación de Empleo, b) Reglamento de prestación de servicios, y c) Proyecto de viabilidad.

Que una vez revisada y analizada la documentación allegada, se vio la necesidad de emitir mediante oficio número SPE-GMSE-2017-EE-0000837 de 06 de abril de 2017, observaciones con relación al Reglamento de Prestación de Servicios, al proyecto de viabilidad y la solicitud de renovación debe estar suscrita por el representante legal de la Universidad.

Que mediante oficio radicado número SPE-SAS-2017-ER-0001128 de 22 de junio de 2017, la **UNIVERSIDAD NACIONAL DE COLOMBIA** a través de su representante legal doctor IGNACIO MANTILLA PRADA, identificado con cédula de ciudadanía No.19328350 de Bogotá, presentó respuesta a las observaciones emitidas.

Que de la documentación allegada, se evidenció que la Bolsa de Gestión y Colocación de Empleo prestará los servicios virtualmente a través del siguiente punto de atención:

a) Punto Virtual Restringido

URL	SERVICIOS
www.egresados.unal.edu.co	Registro de oferentes, demandantes y vacantes Preselección, y Remisión

B. Anexo 2: Código del Open Calais proporcionado por Thomson Reuters.

```
package tr.test;

import org.apache.commons.httpclient.HttpClient;
import org.apache.commons.httpclient.HttpStatus;
import org.apache.commons.httpclient.methods.FileRequestEntity;
import org.apache.commons.httpclient.methods.PostMethod;

import java.io.*;

public class HttpClientCalaisPost {

    private static final String CALAIS_URL = "https://api.thomsonreuters.com/permid/calais";
    public static String uniqueAccessKey;

    private File input;
    private File output;
    private HttpClient client;

    private PostMethod createPostMethod() {

        PostMethod method = new PostMethod(CALAIS_URL);

        // Set mandatory parameters
        method.setRequestHeader("X-AG-Access-Token", uniqueAccessKey);

        method.setRequestHeader("Content-Type", "text/raw");

        // Set response/output format
        method.setRequestHeader("outputformat", "xml/rdf" /*"application/json"*/);

        return method;
    }

    private void run() {
        try {
            if (input.isFile()) {
                postFile(input, createPostMethod());
            } else if (input.isDirectory()) {
                System.out.println("working on all files in " + input.getAbsolutePath());
                for (File file : input.listFiles()) {
                    if (file.isFile())
                        postFile(file, createPostMethod());
                    else

```

```
        System.out.println("skipping "+file.getAbsolutePath());
    }
}
} catch (Exception e) {
    e.printStackTrace();
}
}

private void doRequest(File file, PostMethod method) {
    try {
        int returnCode = client.executeMethod(method);
        if (returnCode == HttpStatus.SC_NOT_IMPLEMENTED) {
            System.err.println("The Post method is not implemented by this URI");
            // still consume the response body
            method.getResponseBodyAsString();
        } else if (returnCode == HttpStatus.SC_OK) {
            System.out.println("File post succeeded: " + file);
            saveResponse(file, method);
        } else {
            System.err.println("File post failed: " + file);
            System.err.println("Got code: " + returnCode);
            System.err.println("response: "+method.getResponseBodyAsString());
        }
    } catch (Exception e) {
        e.printStackTrace();
    } finally {
        method.releaseConnection();
    }
}

private void saveResponse(File file, PostMethod method) throws IOException {
    PrintWriter writer = null;
    try {
        BufferedReader reader = new BufferedReader(new InputStreamReader(
            method.getResponseBodyAsStream(), "UTF-8"));
        File out = new File(output, file.getName() + ".xml");
        writer = new PrintWriter(new BufferedWriter(new FileWriter(out)));
        String line;
        while ((line = reader.readLine()) != null) {
            writer.println(line);
        }
    } catch (IOException e) {
        e.printStackTrace();
    } finally {
        if (writer != null) try {writer.close();} catch (Exception ignored) {}
    }
}

private void postFile(File file, PostMethod method) throws IOException {
```

```

method.setRequestEntity(new FileRequestEntity(file, null));
doRequest(file, method);
}

public static void main(String[] args) {
    verifyArgs(args);
    HttpClientCalaisPost httpClientPost = new HttpClientCalaisPost();
    httpClientPost.input = new File(args[0]);
    httpClientPost.output = new File(args[1]);
    httpClientPost.client = new HttpClient();
    httpClientPost.client.getParams().setParameter("http.useragent", "Calais Rest Client");

    httpClientPost.run();
}

private static void verifyArgs(String[] args) {
    if (args.length==0) {
        usageError("no params supplied");
    } else if (args.length < 3) {
        usageError("3 params are required");
    } else {
        if (!new File(args[0]).exists())
            usageError("file " + args[0] + " doesn't exist");
        File outdir = new File(args[1]);
        if (!outdir.exists() && !outdir.mkdirs())
            usageError("couldn't create output dir");
    }
    uniqueAccessKey = args[2];
}

private static void usageError(String s) {
    System.err.println(s);
    System.err.println("Usage: java " + (new Object() {
}).getClass().getEnclosingClass()).getName() + " input_dir output_dir");
    System.exit(-1);
}
}

```

Bibliografía

- Abello, J. A. D. (2015). Obtener un método para la extracción de información a partir de documentos semiestructurados producidos al interior del Servicio Nacional de Aprendizaje SENA, permitiendo su publicación, reutilización e intercambio a través de la web semántica. Retrieved from <http://www.bdigital.unal.edu.co/51199/1/jhonatanalejandroabellodiaz.2015.pdf>
- Alan Rector et al. (2009). CO-ODE Project | OWL research at the University of Manchester. Retrieved November 22, 2017, from <http://owl.cs.manchester.ac.uk/research/co-ode/>
- Aliprand, J., & Unicode Consortium. (2003). *The Unicode standard*. Addison-Wesley. Retrieved from <https://www.w3.org/TR/rdf-sparql-query/>
- Allard, P., Ferré, S., Allard, P., Ferré, S., Taxonomies, D., Web, S., & Allard, P. (2008). Dynamic Taxonomies for the Semantic Web To cite this version : HAL Id : hal-00284992 Dynamic Taxonomies for the Semantic Web.
- Alter, S. (2012). Metamodel for Service Analysis and Design Based on an Operational View of Service and Service Systems. *Business Analytics and Information Systems.*, 4(3), 218–235. <https://doi.org/10.1287/serv.1120.0020>
- Angles, R., & Gutierrez, C. (2008). The Expressive Power of SPARQL. Retrieved from https://www.dcc.uchile.cl/TR/2006/TR_DCC-2006-009.pdf
- Arenas, M., Cuenca Grau, B., Kharlamov, E., Marciuska, S., & Zheleznyakov, D. (2014). Faceted Search over Ontology-Enhanced RDF Data. *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management - CIKM '14*, 939–948. <https://doi.org/10.1145/2661829.2662027>
- Athanasis, N., Christophides, V., & Kotzinos, D. (2004). Generating On the Fly Queries for the Semantic Web : The ICS-FORTH Graphical RQL Interface (GRQL) 1. *Iswc*, 486–501.
- Bechhofer, S., Harmelen, F. van, Hendler, J., I.Horrocks, McGuinness, D., Patel-Schneider, P., & Stein, L. (2004). OWL Web Ontology Language Reference.

- Retrieved November 22, 2017, from <https://www.w3.org/TR/owl-features/>
- Beneitone, P., Cesar, E., Gonzalez, J., Maleta, M., Siufi, R., & Wagenaar, G. (2004). Reflexiones y perspectivas de la educación superior en América Latina. *Informe Final-Proyecto Tuning—América Latina*, 1–432. Retrieved from <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Reflexiones+y+perspectivas+de+la+Educaci+n+Superior+en+Am+rica+Latina+Informe#1>
- Beneitone, P., González, J., & Wagenaar, R. (ed). (2014). Meta-perfiles y perfiles Una nueva aproximación para las titulaciones en América Latina, 188.
- Boris Villazón-Terrazas, J. R. y A. G.-P. (2014). Human Resources Management Ontology. Retrieved November 20, 2017, from <http://mayor2.dia.fi.upm.es/oeg-upm/index.php/en/ontologies/99-hrmontology/>
- Brunetti, J. M., Garcia, R., & Auer, S. (2013). From Overview to Facets and Pivoting for Interactive Exploration of Semantic Web Data. *International Journal on Semantic Web and Information Systems*, 9(1), 1–20. <https://doi.org/http://dx.doi.org/10.4018/jswis.2013010101>
- Burke, R. (2002). Hybrid Recommender Systems: Survey and Experiments. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.88.8200&rep=rep1&type=pdf>
- Carvalho, Aymamí, & V. (2013). Conceptos Fundamentales y Glosario de Términos del Sistema de Competencia aplicado a la Gestión del Capital Humano. *Consultores Formo Internacional*.
- Congreso de Colombia. (2013). Ley 1636 DE 2013. Retrieved November 6, 2015, from <http://www.alcaldiabogota.gov.co/sisjur/normas/Norma1.jsp?i=53493>
- Consejo Nacional de Acreditación. (2004). CNA - El sistema de educación superior de Colombia. Retrieved February 23, 2018, from <https://www.cna.gov.co/1741/article-187279.html>
- Cunningham, et al. (2017). Developing Language Processing Components with GATE. University of Sheffield Department of Computer Science. Retrieved from <https://gate.ac.uk>
- DAU, F., & Andrews, S. (2013). The CUBIST Project: Combining and Uniting Business Intelligence with Semantic Technologies. *International Journal of Intelligent Information Technologies*, 9(4), 1–15.

- <https://doi.org/http://dx.doi.org/10.4018/ijit.2013100101>
- E. Peis; J. M. Morales-del-Castillo. (2008). *Sistemas de Recomendación Semánticos*. Universidad Pompeu Fabra. Retrieved from <https://www.upf.edu/hipertextnet/numero-6/recomendacion.html>
- Erling, O. (2009). *Faceted Views over Large Scale Linked Data*.
- Ferré, S. (2010). Conceptual navigation in RDF graphs with SPARQL-like queries. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5986 LNAI, 193–208. https://doi.org/10.1007/978-3-642-11928-6_14
- Ferré, S. (2015). Sparklis : a SPARQL Endpoint Explorer for Expressive Question Answering e To cite this version : Sparklis : a SPARQL Endpoint Explorer for Expressive Question Answering, 2–5.
- Francisco Burgos Becerra. (2002). *Formación y orientación laboral: FOL : grado superior, ciclos formativos. Orientaciones y solucionario*.
- González Franco, N., & Alejandres Sánchez, H. O. (2015). *Arquitectura de un sistema de recomendación semántico sensible al contexto para entornos tipo campus*. Retrieved from <http://cinfo.idict.cu/index.php/cinfo/article/viewFile/668/512>
- Hahn, R., Bizer, C., Sahnwaldt, C., Herta, C., Robinson, S., Bürgle, M., ... Scheel, U. (2010). *Faceted Wikipedia search. Lecture Notes in Business Information Processing, 47 LNBIP*, 1–11. https://doi.org/10.1007/978-3-642-12814-1_1
- Harth, A. (2009). *VisiNav: Visual web data search and navigation. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5690 LNCS, 214–228. https://doi.org/10.1007/978-3-642-03573-9_17
- Heim, P., Ertl, T., & Ziegler, J. (2010). *Facet graphs: Complex semantic querying made easy. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 6088 LNCS(PART 1), 288–302. https://doi.org/10.1007/978-3-642-13486-9_20
- Hildebrand, M., Wiskunde, C., Ossenbruggen, J. Van, & Wiskunde, C. (2006). *The Semantic Web - ISWC 2006, 4273(August)*. <https://doi.org/10.1007/11926078>
- Holger Knublauch et al. (2017). *Protégé*. Retrieved from <https://protege.stanford.edu/>
- Holi, M., & Hyvönen, E. (2006). *Fuzzy View-Based Semantic Search*, 351–365. https://doi.org/10.1007/11836025_36

- Horrocks, I., & Patel-Schneider, P. F. (2004). SWRL: A Semantic Web Rule Language Combining OWL and RuleML. Retrieved September 30, 2017, from <https://www.w3.org/Submission/SWRL/>
- Huynh, D., & Karger, D. (2009). Parallax and companion: Set-based browsing for the data web. *WWW Conference, 2005–2008*. Retrieved from <http://davidhuynh.net/media/papers/2009/www2009-parallax.pdf>
- Hyvönen, E., Junnila, M., Kettula, S., Mäkelä, E., Saarela, S., Salminen, M., ... Viljanen, K. (2004). MuseumFinland-Finnish Museums on the Semantic Web User's Perspective Why Museums on the Semantic Web?, 2004. Retrieved from <http://www.cs.helsinki.fi/group/seco/>
- Jaime Carbonell. (1992). CVC. Congreso de Sevilla. La lengua española y las nuevas tecnologías. Retrieved September 24, 2017, from https://cvc.cervantes.es/obref/congresos/sevilla/tecnologias/ponenc_carbonell.htm
- Kobilarov, G., & Dickinson, I. (2008). Humboldt : Exploring Linked Data. *Human Factors*, (April).
- Lei Chen, & Nansheng Yao. (2010). Publishing Linked Data from relational databases using traditional views. In *2010 3rd International Conference on Computer Science and Information Technology* (pp. 9–12). IEEE. <https://doi.org/10.1109/ICCSIT.2010.5563576>
- Loria, C., Gardent, C., & Loria, C. (2015). Parsing Text into RDF graphs. Retrieved from https://gplsi.dlsi.ua.es/sepln15/sites/gplsi.dlsi.ua.es.sepln15/files/attachments/BatoucheEtal_PosterSEPLN2015.pdf
- Maynard, D., Roberts, I., Greenwood, M. A., & Bontcheva, K. (2015). A Framework for Real-time Semantic Social Media Analysis. *Undefined*, 1(1). Retrieved from <http://www.semantic-web-journal.net/system/files/swj1271.pdf>
- Ministerio de Educación Nacional. (2013). Definición de Términos. Retrieved March 16, 2018, from <https://www.mineduacion.gov.co/1621/article-79413.html>
- Musen, M. A. (2015). The protégé project. *AI Matters*, 1(4), 4–12. <https://doi.org/10.1145/2757001.2757003>
- Neo4j Inc. et al. (2017). The Neo4j Graph Platform. Retrieved from <https://neo4j.com/product/?ref=hro>
- Observatorio Laboral para la Educación. (2013). Mapa de Sectores Productivos. Retrieved February 27, 2018, from

- <http://www.graduadoscolombia.edu.co/html/1732/w3-printer-335507.html>
- Oldag, A. (2014). *Joins in SPARQL, Class Notes, Department of Computer Science*. Retrieved from [http://www.cs.utexas.edu/~cannata/cs345/New Class Notes/15 JoinsinSPARQL \(3\).pdf](http://www.cs.utexas.edu/~cannata/cs345/New Class Notes/15 JoinsinSPARQL (3).pdf)
- Oren, E., Delbru, R., & Decker, S. (2006). The Semantic Web - ISWC 2006, 4273(May). <https://doi.org/10.1007/11926078>
- Pietriga. (2006). SIMILE: Longwell RDF Browser.
- Programa de Egresados Universidad Nacional de Colombia. (2009). Situación prospectiva y portafolio de proyectos.
- Quiroga, G. B. B. (2015). Prototipo de un sistema de integración de recursos científicos, diseñado para su funcionamiento en el espacio de los datos abiertos enlazados para mejorar la colaboración, la eficiencia y promover la innovación en Colombia. Retrieved from http://www.bdigital.unal.edu.co/50580/1/MSc_2702295_V50.pdf
- Sarmiento, A., & Baldión, E. (2014). El Desafío de Aumentar la Pertinencia de la Formación Profesional: Propuesta de Metodología para Vincular Profesiones y Ocupaciones y para Identificar la Demanda de Calificaciones por Sector, 54.
- schraefel, m. c., Smith, D. A., Owens, A., Russell, A., Harris, C., & Wilson, M. (2005). The evolving mSpace platform. *Proceedings of the Sixteenth ACM Conference on Hypertext and Hypermedia - HYPERTEXT '05*, (May 2014), 174. <https://doi.org/10.1145/1083356.1083391>
- Tzitzikas, Y., Manolis, N., & Papadakos, P. (2017). Faceted exploration of RDF/S datasets: a survey. *Journal of Intelligent Information Systems*, 48(2), 329–364. <https://doi.org/10.1007/s10844-016-0413-8>
- UNESCO. (2009). Conferencia Mundial sobre la Educación Superior - 2009. *La Nueva Dinámica de La Educación Superior y La Investigación Para El Cambio Social y El Desarrollo*, 1–9. Retrieved from http://www.unesco.org/education/WCHE2009/comunicado_es.pdf
- Universidad Externado de Colombia. (2017). *Boletín del observatorio del mercado del trabajo y la seguridad social*. Retrieved from https://www.uexternado.edu.co/wp-content/uploads/2017/01/Boletin_16.pdf
- van Zwol, R., Ng, P., Ramani, A., Sahai, A., Sathish, S. T., Vasudev, H., ... Murdock, V. (2010). Faceted exploration of image search results. *Proceedings of the 19th International Conference on World Wide Web - WWW '10*, 961.

<https://doi.org/10.1145/1772690.1772788>

World Wide Web Consortium (W3C). (2017). Ontologies - W3C. Retrieved from <https://www.w3.org/standards/semanticweb/ontology>

World Wide Web Consortium et al. (2013). Linked Data Glossary. Retrieved November 22, 2017, from <https://www.w3.org/TR/2013/NOTE-ld-glossary-20130627/>

Yu, L. (2011). *Developer's Guide to the Semantic Web*. (R. V. Brickley, Dan & Guha, Ed.). <https://doi.org/10.1007/978-3-642-15970-1>