

WEBIR



2015

Proyecto CompuJobs - Búsqueda de
ofertas laborales

Integrantes:

Acuña Ignacio
Balduino Mathias
Muñoz Oscar
Serra Diego

Profesor: Libertad Tansini

Introducción

Se pretende realizar un buscador de trabajo unificando ofertas brindadas por grandes buscadores como Computrabajo, Buscojobs entre otros. Para lo cual se realiza una recuperación de los avisos brindados por estos buscadores, brindando luego un buscador más completo y eficiente.

Motivación

Si bien en la actualidad existe una variedad de buscadores de trabajo, éstos están orientados a la búsqueda de avisos realizados sobre la misma aplicación. Por lo tanto, nos resulta interesante poder brindar búsquedas personalizadas, sobre un conjunto amplio de ofertas laborales pudiendo alcanzar el trabajo más adecuado para el solicitante.

Los interesados en publicar avisos de empleo para sus empresas no siempre publican sus avisos en todas los sitios webs, es por esto que se requiere de una aplicación que reúna todas esas ofertas en un sólo sitio.

Además, dada una persona que quiere buscar ofertas de empleo, la persona debe saber a qué página dirigirse para obtener mejores resultados.

Todo esto puede significar que se requiera un aprendizaje de cada sitio web en particular, debido a que todos manejan la información de distinta manera.

Herramientas

Lenguajes

El lenguaje en el cual está implementado el Scrapper es Java.
Para el buscador de ofertas se utilizó HTML y JavaScript.

Recuperación de información

JSoup

Es una librería que permite trabajar con sitios HTML. Además, provee un conjunto de API's para la extracción y manipulación de datos, usando lo mejor del DOM, CSS y jquery como métodos. Se eligió esta herramienta dado que es open source, gratuita, en lenguaje Java y fácil de usar.

Motor de búsqueda

Solr

SOLR es un servidor de búsquedas que trabaja con una API de tipo REST. El objetivo es poner documentos en el (llamado indexación) por medio de JSON, XML, CSV o datos binarios sobre HTTP. Luego se pueden realizar consultas mediante pedidos HTTP GET, para recibir resultados JSON, XML, CSV o binarios. Se eligió esta herramienta dado que es open source, gratuita y fácil de integrar con nuestra solución.

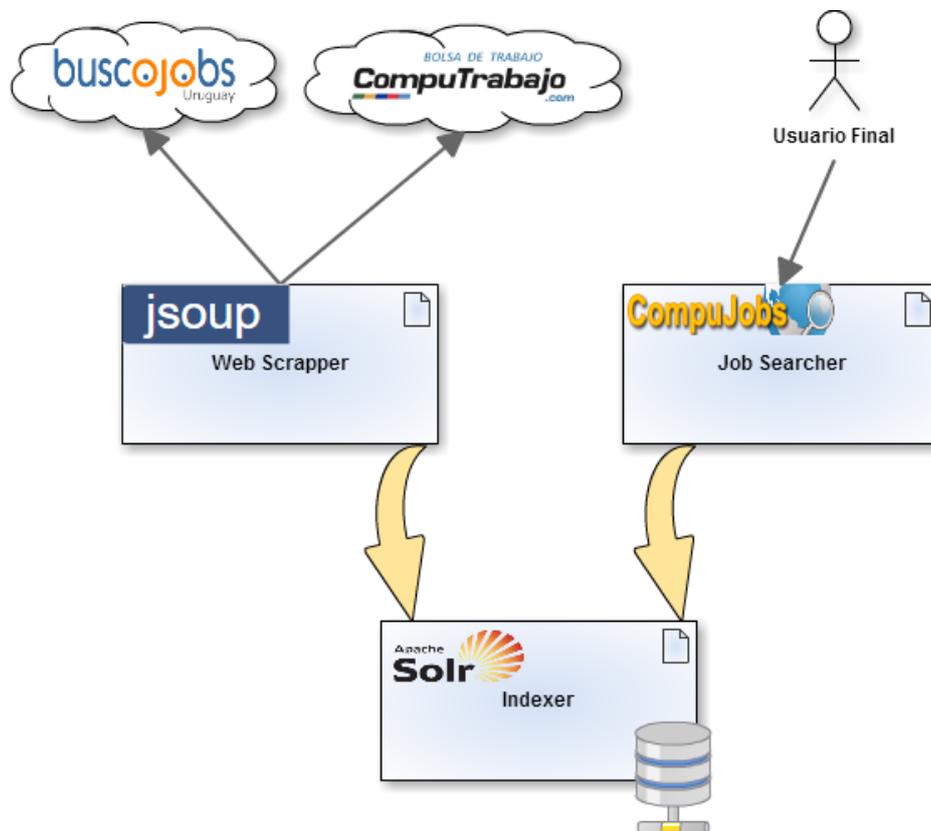
Implementación

Diseño

La aplicación se basa en un programa que, utilizando la técnica de web scrapping mediante la librería JSoup, obtiene información de buscojobs y de CompuTrabajo para luego enviarla al motor de búsqueda Solr.

Solr se encarga de “indexar” los datos que recibe del WebScraper, para que, a la hora de consultarlo la búsqueda sea eficiente. Esta información indexada se almacena en archivos que el mismo Solr sabe manejar.

Por otro lado, se cuenta con la aplicación web CompuJobs, la cual hará consultas a Solr para obtener la información que el usuario requiera.



Web Scraping

La técnica de web scrapping es utilizada para extraer información de sitios web. Los programas que utilizan ésta técnica, generalmente simulan la navegación de un humano utilizando el protocolo HTTP.

Utilizando la técnica de Web Scrapping se implementa un módulo de *Scraper* que realiza la recuperación de los avisos de trabajo actuales desde el sitio web de Computrabajo y el sitio web de BuscoJobs almacenándolos localmente.

Scraper

El módulo scraper se basa en la librería JSoup, como se explicó anteriormente, y mediante los selectores que brinda se obtiene la información de los distintos sitios. Cabe destacar que el módulo está implementado de forma de que se puedan agregar más proveedores de información sin tener un impacto en los demás proveedores, siendo módulos desacoplados.

Sitio CompuJobs

Dado que los lenguajes utilizados para el sistema son puramente *ClientSide*, es posible realizar un deploy del sitio en cualquier servidor Web.

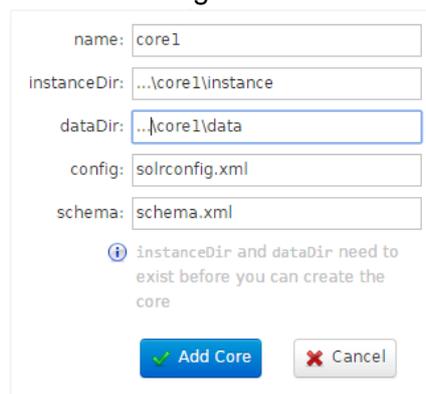
A partir de los datos indexados se implementa el buscador de ofertas de trabajo, para lo cual se generan dos tipos de filtro. El primero de ellos es un filtro genérico, que permite buscar a partir de un conjunto de palabras ingresado en cualquiera de los atributos almacenados, combinando búsquedas sobre los distintos atributos con las palabras ingresadas. Por otro lado se tiene un filtro específico pensado para búsquedas avanzadas sobre los distintos atributos de los avisos.

Manual del usuario.

Instalación

El motor de búsqueda que se utilizó fue Solr, más específicamente la versión 5.3.1. Para su instalación es necesario posicionarse desde la consola (cmd o terminal) en la carpeta bin donde se encuentra Solr y ejecutar el comando “solr start” para iniciar Solr y es levantado en un servidor apache (Jetty) en el puerto 8983 que viene configurado por defecto. Para detenerlo nos posicionamos en el mismo lugar y ejecutamos “solr stop -all”.

Luego de que se levante el motor de búsqueda es necesario crear un “core” el cual tendrá todos los datos indexados además de la estructura de los atributos. En el obligatorio se entregó una carpeta llamada “core1” con la estructura necesaria para generar el nuevo core. Luego se accede al core admin desde la interfaz de usuario de Solr <localhost:8983/solr/#/~cores/> y se crea un nuevo core llamado “core1” llenando los campos con la ubicación de la carpeta core1 del obligatorio.



The image shows a dialog box for adding a new Solr core. It contains the following fields and values:

- name: core1
- instanceDir: ..\core1\instance
- dataDir: ..\core1\data
- config: solrconfig.xml
- schema: schema.xml

Below the fields, there is an information icon and a message: "instanceDir and dataDir need to exist before you can create the core". At the bottom, there are two buttons: "Add Core" (with a green checkmark) and "Cancel" (with a red X).

Una vez hayamos creado el core podemos ejecutar el programa “WebScaper” el cual va a hacer el scraping en las páginas de BuscoJobs y CopuTrabajo para luego indexarlas en nuestro *core*1.

El siguiente y último paso luego de tener los datos indexados es acceder al sitio el cual se comunicará con Solr para realizar las distintas búsquedas que se creen a partir de lo ingresado por el usuario. La primera vez que se entre al sitio web se obtendrán como válidas todas las publicaciones indexadas como muestra en la siguiente imagen, que se podrán ir accediendo mediante un sistema de paginado. Cada vez que se realice una búsqueda el sitio se comunicará con Solr para obtener los resultados de dicha búsqueda.



Búsqueda

Búsqueda Avanzada

Resultados (Encontrados: 2574)

Analista en testing en Montevideo	
Categorías	Tecnologías de la Información
Descripción	El objetivo del puesto es colaborar en la elaboración y ejecución del testing de los diversos proyectos de implementación de sistemas, así como en el análisis de los principales procesos definidos, atendiendo los requerimientos de los clientes. La propuesta se dirige a estudiantes de Ingeniería en sistemas, Analista en Sistemas o similar, que quieran ampliar su experiencia en el rubro, valorándose la experiencia en el mismo o que deseen iniciarse en el mismo. Son competencias requeridas: experiencia en testing y en el área de sistema, el buen relacionamiento interpersonal, el trabajo en equipo y una buena capacidad de análisis. Será valorado el interés y la disposición para aprender sobre el manejo de nuevas tecnologías, procesos y herramientas informáticas. Se trata de una interesante oportunidad para integrarte a un equipo de alto nivel, en una de las mayores firmas de profesionales del país y la región.
Departamento	Montevideo
Empresa	CPA Ferrere

Buscador.

Como ya se mencionó anteriormente, el buscador posee dos tipos de búsquedas, una general y otra avanzada, intentando así lograr que la búsqueda sea lo más eficiente, amigable e intuitivo posible para los usuarios.

La búsqueda general es principalmente un filtro genérico, el cual permite buscar el conjunto de palabras ingresado en cualquiera de los atributos almacenados, logrando encontrar avisos de los cuales algún atributo contenga una parte de lo buscado y las otras palabras buscadas en uno o más atributos.

La búsqueda avanzada se realiza a través de un filtro específico, en el cual se podrán ingresar explícitamente los valores a encontrar en cada atributo del aviso o un rango en caso de valores numéricos, permitiendo componer filtros de búsqueda sobre algunos atributos de la oferta como la descripción, rango de salario, título, años de experiencia mínimos, ubicación, conocimientos requeridos, categoría a la que pertenece.

Búsqueda

Búsqueda Avanzada

Título:	<input type="text"/>	Descripción:	<input type="text"/>
Departamento:	<input type="text"/>	Empresa:	<input type="text"/>
Salario:	<input type="text"/> - <input type="text"/>	Requiere años de experiencia:	<input type="text"/>
Conocimientos Req:	<input type="text"/>	Categorías:	<input type="text"/>

Casos de prueba.

Se realizarán dos casos de prueba para probar los distintos tipos de búsquedas que nos permite el sitio de CompuJobs.

Primero se probará la búsqueda general. Al buscar “Montevideo Java” podrán encontrarse avisos que, por ejemplo, tengan como ubicación Montevideo y requieran conocimientos de Java, o avisos en donde las dos palabras se encuentren en el título, descripción, o cualquier combinación de estas dos palabras en todos los atributos de los avisos indexados. En este caso se encontraron 110 avisos que cumplen con la búsqueda “Montevideo Java”, en el primer resultado podemos observar claramente que la palabra “Montevideo” se encuentra en el título y además en el campo “Departamento” mientras que la palabra “Java” se encuentra en la descripción formando parte de la palabra “Javascript”.



CompuJobs

Búsqueda

Montevideo Java

Búsqueda Avanzada

Resultados (Encontrados: 110)

Referencia 6847: Desarrollador Web SSR. - Montevideo - Uruguay en Montevideo, Montevideo	
Categorías	Diseño / Artes gráficas
Descripción	Referencia 6847: Desarrollador Web SSR. - Montevideo - Uruguay Seniority: Semi Sr Descripción: ADN - Recursos Humanos estamos en la búsqueda de un Desarrollador Web SSR. Montevideo - Uruguay, para importante Empresa Multinacional que provee soluciones tecnológicas para hacer más inteligentes y efectivas las estrategias de marketing digital. Detalle: - Estamos en la búsqueda de un diseñador gráfico con pasión e iniciativa que quiera sumarse al equipo de producto como un miembro clave en la estrategia del negocio. Tendrá como responsabilidad principal el diseño de todos los instrumentos de comunicación e imagen corporativa(internos y externos) tales como sitios web, newsletters, presentaciones comerciales, gráficas para eventos, materiales multimedia, entre otros. También participará en el diseño de la interfaz de la plataforma. La búsqueda está orientada a jóvenes creativos, pragmáticos y responsables, que posean habilidades para manejar diversidad de proyectos en simultáneos de manera autónoma. Se requiere excelente manejo de la estética y usabilidad. Deberá contar con experiencia en diseño web con conocimientos sólidos de armado de newsletters, maquetado html, css, dominio de Photoshop, Illustrator, Dreamweaver, Fireworks, Wordpress y Flash. Conocimiento de Javascript e Inglés serán muy valorado. La experiencia previa en el uso de herramientas de marketing online será un plus. Ofrecemos: * Excelentes condiciones de contratación * Salario acorde a las responsabilidades * Beneficios * Capacitación permanente * Excelente entorno de trabajo y plan de carrera Lugar de Trabajo: Montevideo - Uruguay Lugar de residencia: Preferentemente por la zona. Horario: de Lunes a Viernes de 9 a 18 hs. Edad: Indistinto Sexo: Indistinto Disponibilidad para viajar: - Idioma Inglés: - Fecha de Inicio: Inmediata Modalidad: Relación de Dependencia Duración: a largo plazo Posiciones a cubrir: 1
Departamento	Montevideo
Empresa	ADN - Recursos Humanos
Fecha Publicado	2015-10-28T02:00:00Z
Fecha Expira	-
Url	http://www.uy.computrabajo.com/offeras-de-trabajo/oferta-de-trabajo-de-referencia-6847-desarrollador-web-ssr-montevideo-uruguay-en-montevideo-14C4384A38F8000D
Salario	1
Cant. Vacantes	1
Conocimientos Requeridos	-
Años de experiencia	-

Por último se realizará un caso de prueba de un usuario que realice una búsqueda avanzada. En este caso no se completarán todos los campos de la búsqueda, realizándose sólo con un subconjunto de éstos. Los campos que se completarán serán los de “Categoría” con la palabra “Ingeniería”, “Conocimientos Requeridos” con la palabra “Java” y la búsqueda se realizará para aquellos avisos indexados que posean un salario especificado mayor a \$20.000 y menor a \$300.000, llenando el primer y segundo campo de “salario” con los valores antes mencionados. Se encontró un único aviso indexado que cumple con lo anterior.

Resultados (Encontrados: 1)

Líder de Proyectos en Montevideo	
Categorías	Tecnologías de la Información
Descripción	ABITAB S.A. selecciona para su Area de Tecnología: LÍDERES DE PROYECTO. Estamos orientados a la búsqueda de profesionales con experiencia probada como líderes de proyectos basados en programación JEE y/o Delphi. Requisitos: _ Analista de Sistemas o estudiante universitario avanzado/a en Ingeniería de Sistemas _ Experiencia mínima de 4 años en programación JEE y/o Delphi _ Experiencia en liderar equipos o proyectos. Se valorará _ Experiencia en manejo de Base de Datos, Jboss, JAVA, Webservices y programación Web _ Conocimientos de Design Patterns. Competencias: • Buen relacionamiento interpersonal y vocación por la gestión de personas • Marcada disposición para integrar equipos de trabajo • Orientado a la metodología y calidad en el desarrollo del proyecto Condiciones Laborales: _ Sueldo nominal: entre \$80.000 y \$100.000 mensuales _ 44hs semanales.
Departamento	Montevideo
Empresa	Abitab
Fecha Publicado	2015-11-10T00:47:05.843Z
Fecha Expira	2015-12-10T00:47:05.843Z
Url	http://www.compujobs.com.uy/lder-de-proyectos-ID-119301
Salario	100000
Cant. Vacantes	1
Conocimientos Requeridos	EJB (Enterprise Java Beans) ,J2EE ,Java ,Programación ,Web Services
Años de experiencia	-

Conclusiones

- Único punto de acceso.

Se logró unificar los pedidos de ambas páginas en una sola, mejorando la experiencia de usuario.

- Interfaz sencilla e intuitiva.

Gracias a la interfaz simple que se logró, es más fácil llegar a las ofertas de trabajo que en cualquiera de los sitios de los actuales proveedores.

- Búsquedas eficientes.

Como consecuencia de utilizar un motor de búsqueda como Solr, las búsquedas son eficientes, y se obtiene un muy buen tiempo de respuesta a pesar de la cantidad de datos.

- Filtros de acuerdo a requerimientos de usuarios.

En los sitios actuales, no existen filtros que se adecúen completamente a las necesidades de los usuarios, gracias a CompuJobs, las búsquedas no son sólo un campo de texto, sino que se cuenta con opciones que facilitan la obtención de información por parte de los usuarios.

- Aplicación escalable para nuevos requerimientos.

Toda la aplicación fue pensada para escalar en cantidad de proveedores, es por esto que la única complejidad a la hora de ingresar un nuevo proveedor en el sistema es la extracción de la información en el sitio web del mismo.

- Tiempo de scrapping e indexación aceptables.

Se logró un tiempo de recuperación de información e indexación de aproximadamente 15 minutos por lo que es factible indexar y recolectar información diariamente, esto es importante ya que generalmente estos sitios tienen nuevas publicaciones cada día.

Trabajo a futuro

- Incorporar nuevos sitios de donde extraer la información:

Permitiendo aumentar la cantidad de ofertas de trabajo, permitiendo a los usuarios acceder a más información en un solo lugar.

- Agendar extracción de información en modalidad batch:

Realizando la tarea de extracción de ofertas de trabajo con periodicidad dependiente de la frecuencia con la que se agregan nuevas ofertas de trabajo a cada una de las páginas, por ejemplo podría ser diaria o semanal, permitiendo así tener las últimas ofertas de trabajo de cada sitio.

- Incorporar un módulo de normalización de la información:

Este módulo se encargaría de verificar que los datos extraídos en la base de datos sean consistentes, por ejemplo en el campo salario no deben haber valores negativos.

- Desarrollo de subsistema que permita detectar cambios en las páginas:

El subsistema se encarga de alertar al administrador en caso de que alguna extracción falle, de esta forma se puede actualizar rápidamente el scrapper a partir de lo alertado y no se pierde información.

- Más filtros para cubrir más necesidades:

Al incorporar nuevas páginas de ofertas de trabajo es posible que se incorporen nuevos campos de información, esto implica la inclusión de nuevos tipos de filtros.

- Filtros con combos, integración con solr:

Para los filtros de búsqueda que son combos, es de utilidad que los mismos se obtengan desde el medio de almacenamiento, consultando a Solr. Un claro ejemplo de esto es para el filtro de departamento, es de gran utilidad no permitir seleccionar departamentos para los cuales no hay ofertas de trabajo disponibles.

- Elementos de ayuda en el buscador:

Ofrecer ejemplos o ayuda contextual en los campos del formulario del buscador, para que los usuarios puedan realizar las búsquedas más eficientes según su necesidad.

- Mejoras en la interfaz:

Presentar las ofertas de trabajo en un formato más amigable para el usuario, incluyendo entre otros el logo de la página de donde se extrajo la oferta.

- Incorporar usuarios / Suscripciones a búsquedas:

Permitir realizar búsquedas y suscripciones al resultado de las mismas, al momento de volver a realizar la indexación de los trabajos se verifica si cambiaron los resultados de las búsquedas a las cuales se suscribió el usuario y en ese caso se lo notifica.

- Selectores en archivos de configuración:

Al utilizar archivos de configuración para los selectores no es necesario compilar nuevamente el código fuente del sistema cuando alguna de las páginas de ofertas cambia, de esta forma mejora la mantenibilidad del sistema.

Referencias

[Sitio web de Computrabajo] <http://www.uy.computrabajo.com/>

[Sitio web de Buscojobs] <http://www.buscojobs.com.uy/>

[Jsoup] <http://jsoup.org/>

[Solr] <http://lucene.apache.org/solr/>

[Lucene] <https://lucene.apache.org/>