

BUSCADOR SEMÁNTICO SAWA  
MANUAL DE REFERENCIA

OMAR ERNESTO CABRERA ROSERO  
JIMMY MATEO GUERRERO RESTREPO  
MAURICIO FERNANDO BENAVIDES BENAVIDES  
SILVIO RICARDO TIMARÁN PEREIRA





UNIVERSIDAD DE NARIÑO  
FACULTAD DE INGENIERÍA  
PROGRAMA DE INGENIERÍA DE SISTEMAS  
SAN JUAN DE PASTO  
2014




## Sobre este trabajo



### You are free:

-  to **Share** – to copy, distribute and transmit this work
-  to **Remix** – to adapt this work

### Under the following conditions:

-  **Attribution** – You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work)
-  **Noncommercial** – You may not use this work for commercial purposes.
-  **Share Alike** – If you alter, transform, or build upon this work, you may distribute the resulting work only under the same or similar license to this one.

Subject to conditions outlined in the license.

This work is licensed under the *Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported* License. To view a copy of this license, visit

<http://creativecommons.org/licenses/by-nc-sa/3.0/>

or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, USA.

## CONTENIDO

1. Introducción	4
2. Contrucción de la aplicación	5
3. Funcionamiento de la Aplicación	7
3.0.1. Casos de prueba	7
4. Proceso de Instalación	16

## **1 Introducción**

La falta de significado que se maneja en la Web actual, dificulta la búsqueda eficiente de información. La web semántica ha comenzado a adquirir una gran importancia debido a que se quiere encontrar información de manera precisa y poder convertirla en información del conocimiento y representarla en recursos web que puedan estar disponibles en otras aplicaciones. Para lograr encontrar un significado claro en la búsqueda, la Web semántica hace uso de ontologías, que organizan la información de manera que se podrá interpretar lo que se quiere buscar y, por tanto, permitirá buscar e integrar datos mucho mejor que ahora.

Esta aplicación web soporta la búsqueda inteligente de los trabajos de grado de la Universidad de Nariño mediante una ontología de aplicación.

## 2 Contrucción de la aplicación

El buscador semántico SAWA fue desarrollado en JavaEE<sup>1</sup> y liberado bajo licencia libre GPL3<sup>2</sup> el cual puede ser descargado del repositorio del portal github<sup>3</sup> y se puede mirar el funcionamiento en la página<sup>4</sup>.

En la aplicación se usaron algoritmos como lematizadores y similitud de palabras para hacer corrección ortográfica, en caso de que el usuario tenga error de digitación, además de la construcción de un tesoro para que pueda hacer la búsqueda por sinónimos de palabras.

Las siguientes son las características específicas del software construido:

- Búsqueda General.
- Búsqueda por título.
- Búsqueda por autor.
- Auto completar palabras.
- Corrección de digitación.
- Búsqueda por sinónimos.
- Ordenamiento de resultados por mayor coincidencia.

En el diagrama de actividades de la Figura 2.1 se muestra como se realiza una búsqueda.

Para la construcción del software se tubo en cuenta la manera de realizar las consultas usando el lenguaje SPARQL y la extensión de postgresql pg\_similarity<sup>5</sup>.

---

<sup>1</sup>Java Platform, Enterprise Edition o Java EE (anteriormente conocido como Java 2 Platform, Enterprise Edition o J2EE hasta la versión 1.4

<sup>2</sup><http://www.gnu.org/licenses/gpl.html>

<sup>3</sup><https://github.com/poldrosky/Sawa>

<sup>4</sup><http://ingenieria.udenar.edu.co:8080/Sawa/>

<sup>5</sup><http://pgsimilarity.projects.pgfoundry.org>

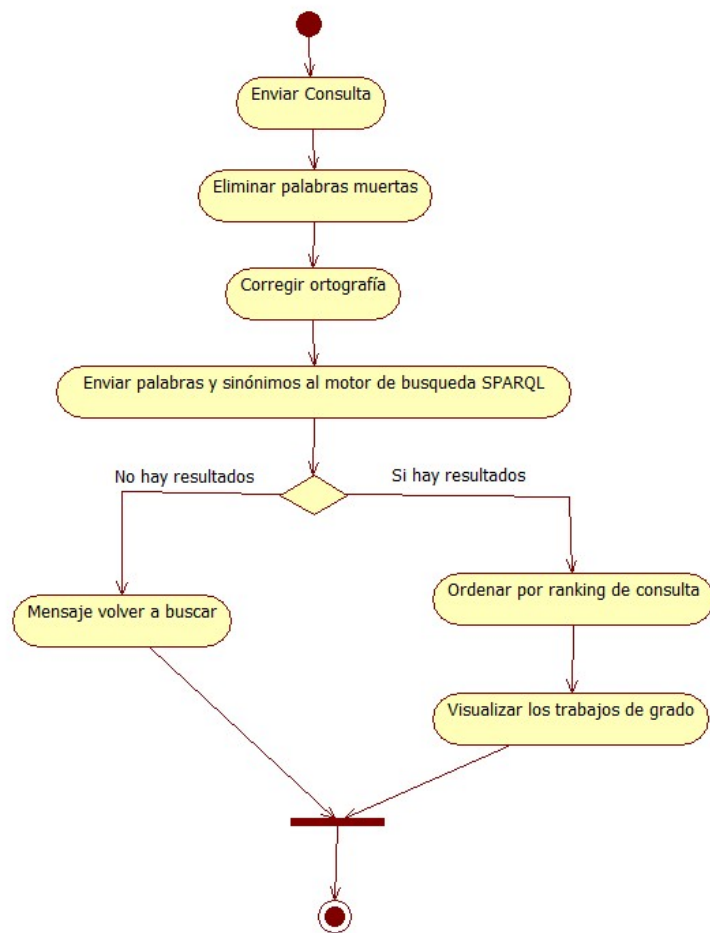


Figura 2.1: Diagrama de actividades de la búsqueda

### 3 Funcionamiento de la Aplicación

Para la aplicación de las pruebas se realizaron siete casos de prueba con diez iteraciones cada una, para ver la eficiencia del buscador desarrollado y el actual buscador de la biblioteca de la Universidad de Nariño<sup>1</sup>.

Las pruebas se hicieron llevando los siguientes casos de prueba y se calificó como éxito (E) o fracaso (F), teniendo en cuenta que el éxito se lo califica si la búsqueda a realizar está en los quince primeros resultados.

#### 3.0.1. Casos de prueba

**Búsqueda por título completo:** esta búsqueda se la realizó enviando la consulta con el título exacto que aparece en la base de datos(tildes, signos de puntuación y comillas), dando como resultado que los dos sistemas se comportan de la misma manera como lo muestra la Tabla 3.1 y Figura 3.1.

---

<sup>1</sup><http://biblioteca.udenar.edu.co:8085/bibliotecavirtual/>

Tabla 3.1: Búsqueda por título completo

No	Consulta	SAWA	Biblioteca
1	MATE-KDD: UNA HERRAMIENTA GENERICA PARA EL DESCUBRIMIENTO DE REGLAS DE CLASIFICACION MEDIANAMENTE ACOPLADA AL SGBD POSTGRESQL	E	E
2	TARIYKDD : UNA HERRAMIENTA GENERICA DE DESCUBRIMIENTO DE CONOCIMIENTO EN BASES DE DATOS DEBILMENTE ACOPLADA CON EL SGBD POSTGRESQL	E	E
3	IMPLANTACION DE PRIMITIVAS SQL PARA EL DESCUBRIMIENTO DE REGLAS DE ASOCIACION Y CLASIFICACION	E	E
4	ATLAS HERRAMIENTA DE CARTOGRAFIA WEB Y GEOCODIFICACION PARA EL DESARROLLO DE SISTEMAS HIBRIDOS EN AREAS URBANAS SOBRE J2EE Y POSTGRESQL	E	E
5	GEOPASTO UN SISTEMA DE INFORMACION GEOGRAFICA WEB ORIENTADO AL APOYO PARA LA TOMA DE DECISIONES BASADOS EN EL PLAN DE ORDENAMIENTO TERRITORIAL	E	E
6	SISTEMA DE INFORMACION PARA EL REGISTRO Y CONSULTA DEL MATERIAL DE BIBLIOTECA EN EL ENTORNO INTRANET DEL CENTRO SUR COLOMBIANO DE LOGISTICA INTERNACIONAL	E	E
7	SISTEMA DE INFORMACION PARA LA BIBLIOTECA Y DESARROLLO DE LA PAGINA WEB DE LA UNIVERSIDAD DE NARIÑO -SEDE IPIALES	E	E
8	ANALISIS, DISEÑO Y DESARROLLO DE UN SISTEMA DE INFORMACION MEDIANTE UNA RED DE COMUNICACIONES E INTERNET PARA LA ALCALDIA DE LINARES.	E	E
9	SISTEMA DE INFORMACION PARA EL MANEJO CONTABLE DE LOS ALMACENES DEL GRUPO CABAL IPIALES	E	E
10	DESARROLLO DE UN SISTEMA DE INFORMACION COMPUTARIZADO DE REGISTRO Y CONTROL ACADEMICO PARA EL POLITECNICO SAN JUAN DE PASTO	E	E

**Búsqueda por autor completo:** esta búsqueda se la realizó enviando la consulta con el autor exacto que aparece en la base de datos, dando como resultado que el sistema realizado en esta investigación de un resultado del 100 % y el sistema de biblioteca 40 % como lo muestra la Tabla 3.2 y Figura 3.2.

Tabla 3.2: Búsqueda por autor completo

No	Consulta	SAWA	Biblioteca
1	CLAUDIA MILENA CASTRO RODRIGUEZ	E	F
2	ANDRES OSWALDO CALDERON ROMERO	E	E
3	STIVENSON ARMERO KREISBERGER	E	E
4	CARLOS ERNESTO ARTEAGA NOGUERA	E	F
5	JUAN CARLOS ROMAN FIGUEROA	E	E
6	HECTOR EDMUNDO ROSERO CASTRO	E	F
7	PAOLA MARY CORAL BASTIDAS	E	F
8	OCTAVIO DELGADO ORDOÑEZ	E	E
9	MARIA ELENA ESTRADA ESPAÑA	E	F
10	YEIMY ANDRES ARTEAGA GUERRON	E	F



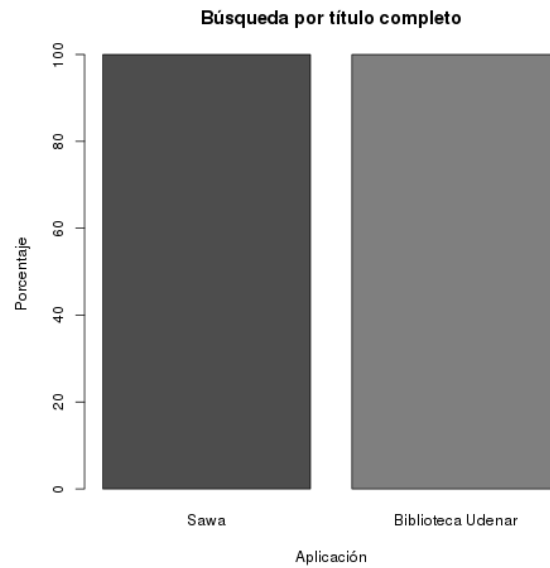


Figura 3.1: Búsqueda por título completo

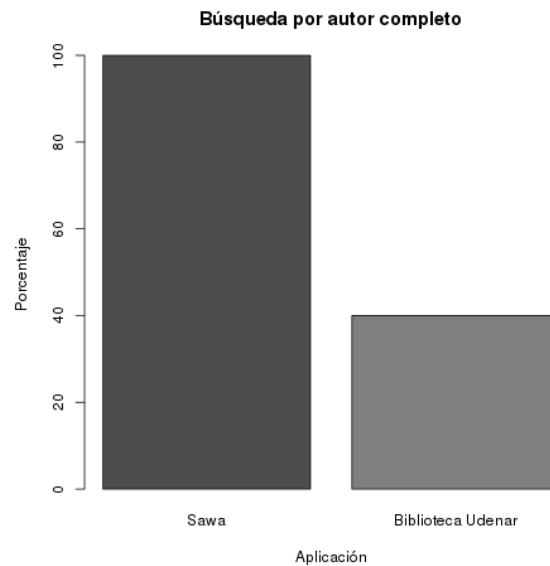


Figura 3.2: Búsqueda por autor completo

**Búsqueda por un nombre y un apellido:** esta búsqueda se la realizó enviando la consulta con un nombre y un autor unicamente, dando como resultado que el sistema realizado en esta investigación de un resultado del 90 % y el sistema de

biblioteca 40 % como lo muestra la Tabla 3.3 y Figura 3.3.

Tabla 3.3: Búsqueda por un nombre y un apellido

No	Consulta	SAWA	Biblioteca
1	CLAUDIA RODRIGUEZ	E	F
2	ANDRES CALDERON	E	E
3	STIVENSON KREISBERGER	E	E
4	CARLOS NOGUERA	E	F
5	JUAN FIGUEROA	E	E
6	HECTOR ROSERO	E	F
7	MARY CORAL	E	F
8	OCTAVIO DELGADO	E	E
9	ELENA ESPAÑA	F	F
10	YEIMY ARTEAGA	E	F

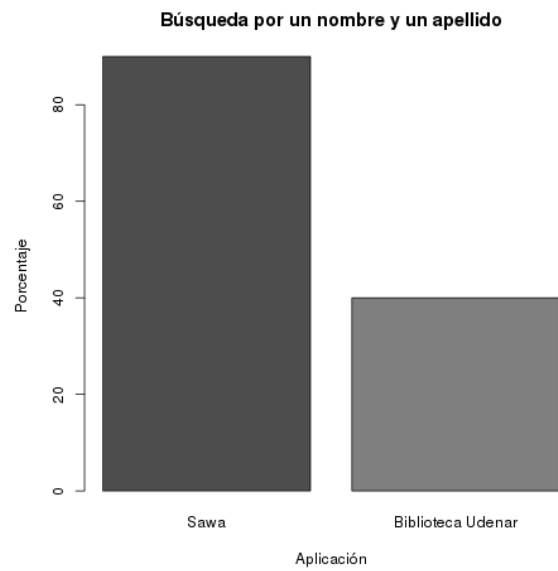


Figura 3.3: Búsqueda por un nombre y un apellido

**Búsqueda por palabras contenidas en el título:** esta búsqueda se la realizó enviando la consulta con palabras que estén contenidas en el título, las palabras podían haberse enviado en el orden del título como en desorden, dando como resultado que el sistema realizado en esta investigación de un resultado del 100 % y el sistema de biblioteca 0 % como lo muestra la Tabla 3.4 y Figura 3.4.

Tabla 3.4: Búsqueda por palabras contenidas en el título

No	Consulta	SAWA	Biblioteca
1	REGLAS GENERICAS DE CLASIFICACION	E	F
2	CONOCIMIENTO ACOPLADO A UNA BASE DE DATOS	E	F
3	PRIMITIVAS PARA EL DESCUBRIMIENTO DE REGLAS	E	F
4	CARTOGRAFIA Y GEOCODIFICACION WEB	E	F
5	GEOPASTO ORIENTADO ALA TOMA DE DECISIONES	E	F
6	REGISTRO DEL MATERIAL DE BIBLIOTECA	E	F
7	DESARROLLO DEL SISTEMA PARA LA BIBLIOTECA	E	F
8	ANALISIS Y DESARROLLO DE UNA UNA RED DE COMUNICACIONES	E	F
9	SISTEMA PARA EL MANEJO CONTABLE DEL GRUPO CABAL	E	F
10	SISTEMA DE INFORMACION DE REGISTRO Y CONTROL ACADEMICO	E	F

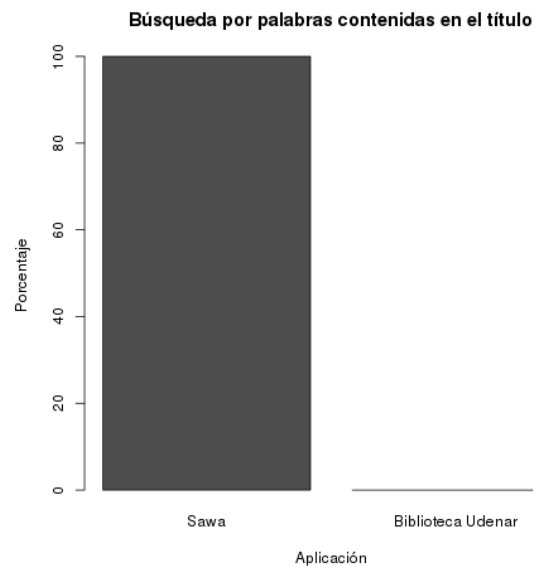


Figura 3.4: Búsqueda por palabras contenidas en el título

**Búsqueda por error ortográfico en el título:** esta búsqueda se la realizó enviando la consulta con uno o dos errores de digitación en el título, dando como resultado que el sistema realizado en esta investigación de un resultado del 100 % y el sistema de biblioteca 0 % como lo muestra la Tabla 3.5 y Figura 3.5.

Tabla 3.5: Búsqueda por error ortográfico en el título

No	Consulta	SAWA	Biblioteca
1	MATE-KDD UNA HERRAMIENTA GENERICA PARA EL DESCUBRIMIENTO DE REGLOS DE CLASIFICACION MEDIANAMENTE ACOPLADA AL SGBD POSTGRESQL	E	F
2	TARIYKDD UNA HERRAMIENTA GENERICA DE DESCUBRIMIENTO DE CONOCIMIENTO EN BASES DE DATOS DEBILMENTE ACOPLADA CON EL SGBD POSTGRESQL	E	F
3	IMPLANTACION DE PRIMITIVAS SQL PARA EL DESCUBRIMINTO DE REGLAS DE ASOSIASION Y CLASIFICACION	E	F
4	ATLAS HERRAMIENTA DE CARTOGAFIA WEB Y GEOCODIFICACION PARA EL DESARROLLO DE SISTEMAS HIBRIDOS EN AREAS URBANAS SOBRE JEE Y POSTGRES	E	F
5	GEOPATO UN SISTEMA DE INFORMACION GEOGAFICA WEB ORIENTADO AL APOYO PARA LA TOMA DE DECISIONES BASADOS EN EL PLAN DE ORDENAMIENTO TERRITORIAL	E	F
6	SISTEMA DE INFORMACIOM PARA EL REGISTRO Y CONSULT DEL MATERIAL DE BIBLIOTECA EN EL ENTORNO INTRANET DEL CENTRO SUR COLOMBIANO DE LOGISTICA INTERNACIONAL	E	F
7	SISTEMA DE INFORMACION PARA LA BIBLIOTECA Y DESARROLLO DE LA PAGINA WEB DE LA UNIVERSIDAD DE NARINO IPIALES	E	F
8	ANALISIS DISENO Y DESARROLLO DE UN SISTEMA DE INFORMACION MEDIANTE UNA RED DE COMUNICACIONES E INTERNET PARA LA ALCALDIA DE LINARES.	E	F
9	SISTEMA DE INFORMACION PARA EL MANEJO CONTABE DE LOS ALMASENES DEL GRUPO CABAL IPIALES	E	F
10	DESARROLLO DE UN SISTEMA DE INFORMACION COMPUTALIZADO DE REGISTRO Y CONTROL ACADEMIC PARA EL POLITENICO SAN JUAN DE PASTO	E	F

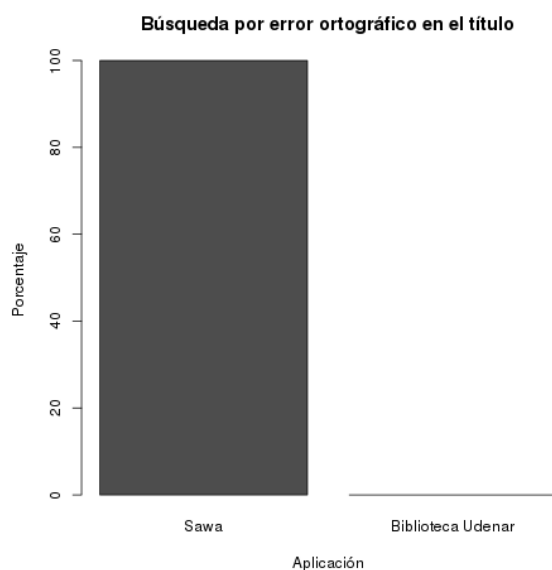


Figura 3.5: Búsqueda por error ortográfico en el título

**Búsqueda por error ortográfico en el autor:** esta búsqueda se la realizó enviando la consulta con uno o dos errores de digitación en el auto, dando como resultado que el sistema realizado en esta investigación de un resultado del 100 % y el sistema de biblioteca 0 % como lo muestra la Tabla 3.6 y Figura 3.6.

Tabla 3.6: Búsqueda por error ortográfico en el autor

No	Consulta	SAWA	Biblioteca
1	CLAUDA MILENA CASTRO RODRIGUEZ	E	F
2	ANDES OSWALDO CALDEROM ROMERO	E	F
3	STIVENSOM ARMERA KREISBERGER	E	F
4	CARLOS ERNESSTO ATEAGA NOGUERA	E	F
5	JUAN CARLS ROMAM FIGUEROA	E	F
6	ECTOR EDMUDO ROSERO CASTRO	E	F
7	PAOA MARYS CORAL BASTIDAS	E	F
8	OCTAIO DELGADO ORDOÑEZ	E	F
9	MARA ELENA ESTADA ESPAÑA	E	F
10	YEIMMY ANDRESS ARTEAGA GUERRON	E	F

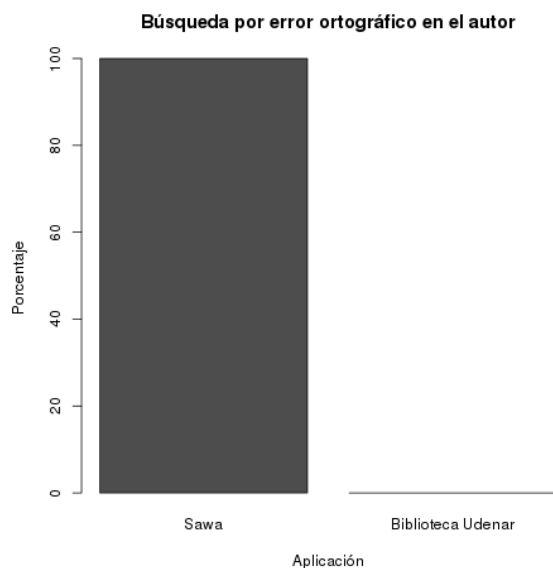


Figura 3.6: Búsqueda por error ortográfico en el autor

**Búsqueda por sinónimos de palabras:** esta búsqueda se la realizó enviando la consulta con algunos de los sinónimos de palabras contenidas en el título, dando como resultado que el sistema realizado en esta investigación de un resultado del 80 % y el sistema de biblioteca 0 % como lo muestra la Tabla 3.7 y Figura 3.7.

Tabla 3.7: Búsqueda por sinónimos de palabras

No	Consulta	SAWA	Biblioteca
1	entendimiento adaptado a una base bases datos	E	F
2	intrumento de hallazgo para conocer	E	F
3	elemental hallazgo DE REGLAS	E	F
4	georreferenciación y localización web	E	F
5	geopasto poner tomas de determinación	E	F
6	inspección e informe de archivo o biblioteca	E	F
7	creacion del sistema para archivo	E	F
8	estudio y creacion de un a red de comunicado	F	F
9	empleo de la asociación cabal	F	F
10	desarrollo de un sistema de inspección para el control escolar	E	F

Se puede mirar el funcionamiento del software y como fueron ejecutadas las pruebas en el video adjunto (“Documentacion/Pruebas.mp4”) o en línea<sup>2</sup>.

<sup>2</sup><http://www.youtube.com/watch?v=niE-FcWzp6Q>

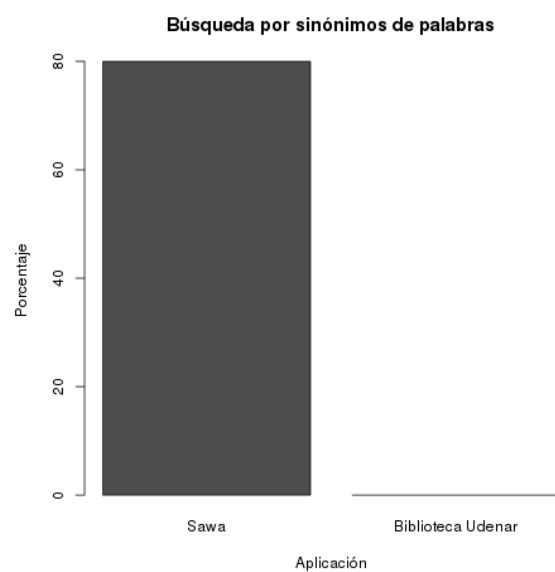


Figura 3.7: Búsqueda por sinónimos de palabras

## 4 Proceso de Instalación

Para poder instalar la aplicación debemos instalar PostgreSQL 9.1, Glassfish 3.1.2.2 y la extensión para postgresql llamada pg\_similarity, la aplicación solo puede ser instalada bajo sistemas operativos GNU/Linux

Para la instalación de PostgreSQL 9.1 (Para Debian/Linux), en una terminal se ejecuta los siguiente:

```
$ su
# apt-get install postgresql postgresql-client pgadmin3
postgresql-server-dev-all postgresql-contrib
```

Ahora se borra la contraseña para la cuenta de administrador de “postgres” para ello se ejecuta lo siguiente en la linea de comandos.

```
# su postgres -c psql template1
```

```
template1=# ALTER USER postgres WITH PASSWORD ‘postgres1’;
template1=#\q
```

Eso altera la contraseña dentro de la base de datos, ahora se tiene que hacer lo mismo para el usuario ‘postgres’ y colocar la misma contraseña que utilizó anteriormente.

```
# passwd -d postgres
#su postgres -c passwd
```

Ahora para instalar pg\_similarity primero se descarga desde github de la siguiente manera

```
$ git clone https://github.com/eulerto/pg_similarity.git
```

Luego para que pueda ser usada la extensión se la compila como superusuario de la siguiente manera.



```
# su
# cd pg_similarity
# USE_PGXS=1 make
# USE_PGXS=1 make install
# su postgres
$ createdb sawa
```

Luego se crea las funciones y se restaura la copia de la base de datos sawa.sql

```
$ psql sawa
sawa=# create extension pg_similarity;
sawa=# \i sawa.sql
```

## Instalación de Glassfish

Primero se descarga la versión 3.1.2.2 para GNU/Linux desde la página de Oracle<sup>1</sup> y se la ejecuta.

```
$ sh ogs-3.1.2.2-unix-ml.sh
```

Hay que descargar el driver JDBC de postgresql desde la página de PostgreSQL<sup>2</sup> y copiarlo en el directorio glassfish3/glassfish/domains/domain1/lib. Para iniciar el servidor se ejecuta lo siguiente

```
$ ./glassfish3/glassfish/bin/startserv
```

Con esto ya en el navegador se ingresa con la dirección <http://localhost:8080>, y entramos a la consola de administración, ingresamos usuario y contraseña de haber escrito una en la instalación de glassfish.

Ir a “Resources/JDBC/Connection Pools” y crear una nueva conexión con los datos que muestra la Figura 4.1 y luego clic en siguiente.

Seleccione el origen de datos de nombre de clase `org.postgresql.ds.PGConnectionPoolDataSource` y escribir a las siguientes propiedades adicionales como muestra la Figura 4.2.

Ya con esto guardamos las conexiones y damos clic en finalizar para guardar la conexión.

---

<sup>1</sup><http://www.oracle.com/technetwork/middleware/glassfish/downloads/ogs-3-1-1-downloads-439803.html>

<sup>2</sup><http://jdbc.postgresql.org>

**New JDBC Connection Pool (Step 1 of 2)**  
Identify the general settings for the connection pool.

**General Settings**

**Pool Name:**

**Resource Type:**   
Must be specified if the datasource class implements more than 1 of the interface.

**Database Driver Vendor:**   
Select or enter a database driver vendor

**Inspect:** ☐ **Enabled**  
If enabled, data source or driver implementation class names will enable introspection.

Figura 4.1: Crear Nueva conexión

Name	Value
portNumber	5432
databaseName	sawa
dataSourceName	jdbc/sawa
roleName	
networkProtocol	
serverName	localhost
user	postgres
password	postgres1

Figura 4.2: Propiedades adicionales de conexión

Luego en “Resources/JDBC/JDBC Resources” escribimos en nombre JNDI y escogemos el Pool Name creado anteriormente como lo muestra la Figura 4.3.

Ya por último en “Applications” escogemos en donde tenemos almacenada la aplicación web que tiene extensión “.war” como muestra la Figura 4.4.

Si todo sale bien nos aparecerá una ventana con la url de la aplicación como muestra la Figura 4.5. y al ingresar a la aplicación nos desplegara la ventana de inicio de la aplicación como lo muestra la Figura 4.6.

**Nota:** Si no desea instalar la aplicación y quiere ver su funcionamiento la aplicación esta alojada en <http://ingenieria.udenar.edu.co:8080/Sawa/>

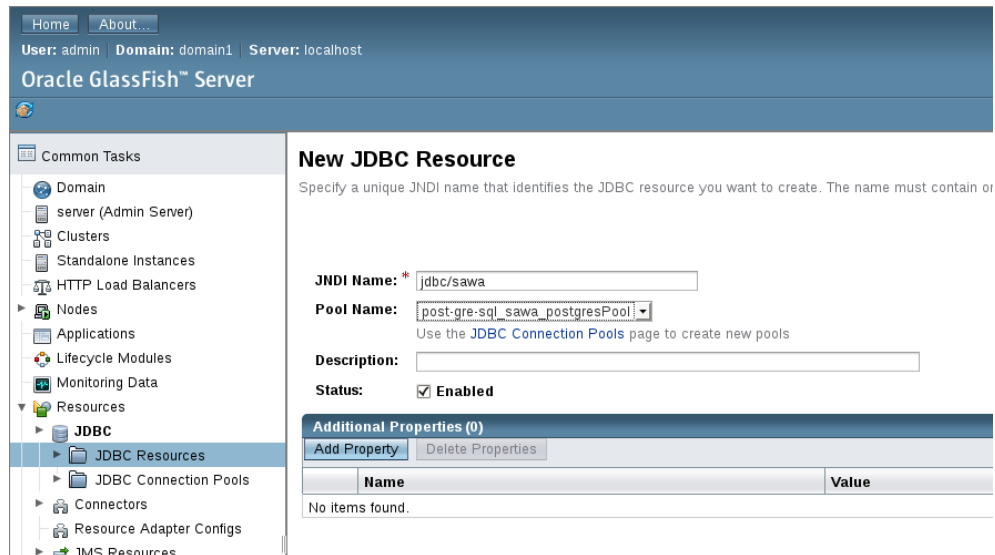


Figura 4.3: Recursos JDBC

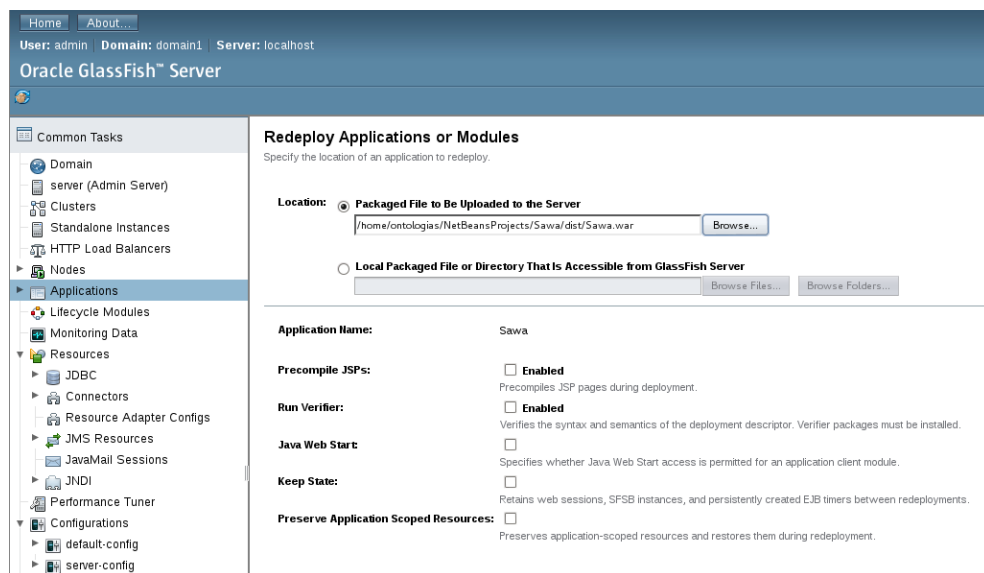


Figura 4.4: Subir Aplicación

## Web Application Links

If the server or listener is not running, the link may not work. In this event, check

**Application Name:** Sawa

**Links:**  
[server] <http://grias:8080/Sawa>  
[server] <https://grias:8181/Sawa>

Figura 4.5: Url de la Aplicación



The screenshot displays the web interface of the 'Buscador Semántico SAVA' application. At the top, a banner features the University of Nariño logo and the text 'Universidad de Nariño' with the tagline 'Entregamos Futuro con Calidad'. Below the banner, the left sidebar contains the university's official seal and lists the authors (Omar Cabrera, Mateo Guerrero, Mauricio Benavides) and the director (Ricardo Timarín Pereira, PhD). The main content area is titled 'Buscador Semántico SAVA' and includes three tabs: 'GENERAL' (selected), 'TITULO', and 'AUTOR'. A search input field labeled 'Buscar' is positioned below the tabs, with a 'Buscar' button to its right. The footer of the application identifies the 'Grupo de Investigación Aplicada en Sistemas (GRIAS)' at the 'Universidad de Nariño'.

Figura 4.6: Aplicación