

LINKED DATA APLICADO A LA EDUCACIÓN: ESTUDIANTES DE NIVEL SUPERIOR INSCRITOS EN CURSOS DENTRO DE LA UNIVERSIDAD OPEN UNIVERSITY

COMPONENTE:

SISTEMAS BASADOS EN EL CONOCIMIENTO

AUTORES:

VANESSA KATHERINE SOTOMAYOR ARMIJOS JUAN PABLO ARROBO LAZO

OCT2018-FEB2019

LOJA- ECUADOR

Contenido

LINKED DATA APLICADO A LA EDUCACIÓN: ESTUDIANTES DE NIVEL SUPERIOR INSCRITOS CURSOS DENTRO DE LA UNIVERSIDAD OPEN UNIVERSITY	
1. INTRODUCCIÓN:	
3. OBJETIVOS Y BENEFICIARIOS:	
4. METODOLOGÍA	
5. ESPECIFICACIÓN DE FUENTES DE DATOS	5
6. ESPECIFICACIÓN, CONCEPTUALIZACIÓN Y FORMALIZACIÓN DEL MODELO ONTOLÓGICO	6
a. Propósito	6
b. Alcance	6
c. Lenguaje de Programación	б
d. Posibles usuarios finales	6
e. Posibles usos	6
f. Requerimientos	б
g. Glosario de términos	7
h. Diagrama de relaciones binarias	7
i. Diccionario de clases	٤
j. Matriz de atributos	8
k. Tabla de Relaciones Binarias	8
I. Modelo formal de la ontología	9
7. DESARROLLO	10
7.1. Análisis y Desarrollo de Ontología	10
7.2. Refinamiento y tratamiento de Datos	10
7.3. Publicación de la ontología en Virtuoso con Maquita Virtual en Digital Ocean	s . 10
7.4. Explotación y desarrollo de Aplicación para consulta de Datos	12
8 BIBLIOGRAFÍA	17

LINKED DATA APLICADO A LA EDUCACIÓN: ESTUDIANTES DE NIVEL SUPERIOR INSCRITOS EN CURSOS DENTRO DE LA UNIVERSIDAD OPEN UNIVERSITY

Vanessa Sotomayor¹, Juan Pablo Arrobo²

Departamento de Tecnologías avanzadas de la Web

Universidad Técnica Particular de Loja

San Cayetano Alto, Loja - Ecuador

¹vksotomayor@utpl.edu.ec, ² jparrobox@utpl.edu.ec

Abstract. Las representaciones en que se basa la web actual básicamente describen la forma de mostrar la información, pero no su significado. La tecnología básica para la representación del conocimiento en la Web Semántica es la ontología.

La Web Semántica, como una nueva visión, propone añadir un componente descriptivo a los recursos disponibles, que extienda las representaciones vigentes. En el presente trabajo, se expone un estudio sobre gestión de sitios, web semántica y sus tecnologías, como base para una propuesta de descripción de un sitio web, como un proceso más natural que tenga en cuenta la semántica del contenido gestionado, El tipo de aplicación se ha definido principalmente por cómo el conocimiento se ha representado y difundido.

1. INTRODUCCIÓN:

La web Semántica creada por TIM BERNERS LEE, que con ella lo que se busca es que las maquinas puedan entender y utilizar la información que se encuentra colgada en la WEB tradicional, y de esta manera poder realizar búsquedas mas precisas e inteligentes dentro de la web, esto nos lleva a tener una evolución de la web y que esta navegación se lleve a cabo a través de datos estructurados y enlazados. Hoy en día podemos ver un gran crecimiento en la nube de Linked Data con lo que se ha desarrollado nuevas herramientas y técnicas para la publicación y consumo de Datos enlazados.

2. JUSTIFICACIÓN:

Con la idea principal de publicar conjuntos de datos abiertos y entrelazados bajo un contexto de Web Semántica en la que se pretende dar una solución a la web actual, se pretende estructurar semánticamente la información relaciona con la educación superior, Para ello hemos utilizado diversas tecnologías como son RFD (lenguaje estándar para la publicación de datos semánticos), SPALQL (lenguaje recomendado por la W3C para la consulta de datos RDF), VIRTUOSO OPEN SOURCE (herramienta en la cual se almacena datos RDF), OPEN RIFINE (herramienta para la limpieza de datos), CMAPS TOOLS (para la conceptualización de la información), PROTEGE (herramienta para la creación de vocabularios nuevos)

El dataset utilizado para este proyecto servirá como referencia para unificar la información que generan los estudiantes al interactuar con un curso dentro de un entorno de aprendizaje, se puede evidenciar una gran cantidad de datos y recursos académicos aislados que carecen de significado, por lo que es de gran importancia buscar una solución de accesibilidad a los datos académicos y educativos según principios establecidos por LINKED DATA para vincular datos en base a su significado facilitándonos la búsqueda.

3. OBJETIVOS Y BENEFICIARIOS:

- Investigar los principios y tecnologías Linked Data, y su aplicación en el ámbito educativo (Superior)
- Estructurar los datos académicos, estandarizándolos a un formato de datos RDF para su posterior publicación.
- Publicar los datos educativos estandarizados, siguiendo los principios y tecnologías Linked Data.
- Como grupo beneficiario se tienen a estudiantes que deseen inscribirse en un curso dentro de un entorno virtual de aprendizaje.

4. METODOLOGÍA

¿Como realizar?: Se hace un estudio breve sobre la web semántica y las tecnologías Linked Data: RDF, RDFS, SPARQL; y de las herramientas que facilitan la publicación y generación de datos enlazados. También se hace referencia a algunos de los proyectos linked data relacionados al ámbito educativo

Materiales:

Los principales materiales que se utilizó son: un computador portátil Lenovo T470 con Sistema Operativo Windows 10, en el que se configuran las herramientas necesarias para ejecución de cada una de las etapas de la metodología de generación y publicación de datos enlazados, como son:

- Protégé,
- CMAPSTOOLS,
- Open Refine
- Servidor Open Source,
- Sparql
- Angular 6
- Virtuoso

Método Científico:

El método científico nos permitió, obtener la información necesaria para analizar y seleccionar las herramientas aplicadas en cada una de las fases de la presente investigación.

Método Deductivo:

Este método se utilizó para la recolección de información y definición del objetivo general, además se logró poner en práctica: definiciones relacionadas con Web Semántica y Linked Data. El método deductivo se utilizará en la fase de Identificación de la Fuente de Datos, para poder: determinar la situación actual sobre la estructuración de la información educativa.

TÉCNICAS DE RECOLECCIÓN DE INFORMACIÓN

Observación.

La técnica de la observación permitió conocer la forma en que se encuentra estructurada la información que se consume en la web referente a la educación superior.

Investigación Bibliográfica:

Esta técnica permite sustentar la parte teórica de la investigación, mediante consultas a: bases de datos científicas, bibliotecas virtuales, tesis, artículos científicos, posters y otras fuentes informáticas.

CICLO DE PUBLICACIÓN DE DATOS ENLAZADOS.



5. ESPECIFICACIÓN DE FUENTES DE DATOS

Este conjunto de datos ofrece dos de los elementos en el marco: comportamiento y rendimiento. Contiene información sobre 22 cursos, 32,593 estudiantes, sus resultados de evaluación y registros de sus interacciones con el VLE representados por resúmenes diarios de clics de estudiantes (10,655,280 entradas).

6. ESPECIFICACIÓN, CONCEPTUALIZACIÓN Y FORMALIZACIÓN DEL MODELO ONTOLÓGICO

a. Propósito

La ontología sobre el Sistema de Educación a través de VLE de la Universidad Open University para describir los términos asociados al dominio indicado.

b. Alcance

La ontología sobre el Sistema de Educación a través de VLE de la Universidad Open University para un uso genera. El nivel de granularidad está directamente relacionado con las cuestiones de competencia que se especifican después y los términos identificados.

c. Lenguaje de Programación

La ontología será implementada en el lenguaje OWL utilizando la herramienta de edición de ontologías Protegé.

d. Posibles usuarios finales

- > Investigadores que deseen tener referencia sobre VLE en el mundo
- Estudiantes interesados en seguir un curso a través de VLE.

e. Posibles usos

- > Buscar información sobre cursos en VLE.
- Universidades o instituciones que quieran ofrecer cursos en VLE.

f. Requerimientos

> Requerimientos no funcionales

La ontología se desarrollará completamente en idioma inglés.

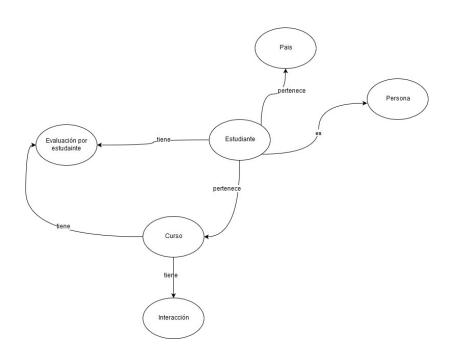
> Requerimientos funcionales

- 1. Conocer la cantidad de estudiantes matriculados en un curso
- 2. Conocer estudiante con mayor puntaje en un curso.
- 3. Conocer a qué país pertenecen la mayoría de los estudiantes registrados
- 4. Conocer los cursos con mayor cantidad de estudiantes
- 5. Conocer el número de interacciones del estudiante a través de la plataforma
- 6. ¿Cuáles son los estudiantes con mayor puntaje en X curso?
- 7. ¿Qué tipo de evaluación tiene mayor puntaje?
- 8. ¿Qué curso tiene mayor cantidad de estudiantes?
- 9. ¿A qué genero pertenecen la mayoría de los estudiantes?
- 10. ¿Cuántas evaluaciones tiene un curso?

g. Glosario de términos

NOMBRE	SINÓNIMO	ACRÓNIMOS	DESCRIPCIÓN	TIPO
Estudiante			Cualquier persona que se inscriba a un VLE	Concepto
Curso			Especialidad de formación en un semestre	Concepto
Interacción			Veces que el estudiante interactúa con el curso	Concepto
Evaluación Estudiante			Evaluación del estudiante sobre el curso	Concepto
País			Región al que el Estudiante pertenece al cursar el modulo	Concepto
Tiene			El concepto principal de la entidad que prosigue	Relación
Pertenece				Relación
Nivel de educación			Nivel de instrucción del estudiante al tomar el curso.	Atributo
Identificador Estudiante			Identificador asignado al estudiante al momento del registro.	Atributo
Género			Característica diferenciada que permite conocer si es hombre o mujer.	Atributo
Código Modulo			Código de identificación para un módulo	Atributo
Código Presentación			Código de identificación de la presentación.	Atributo
Nombre País			Identifica el nombre región geográfica, donde vivió el estudiante mientras tomaba el módulo de presentación.	Atributo
Cantidad Interacciones			La cantidad de veces que un estudiante interactúa con el material en ese día.	Atributo
Fecha Interacciones			La fecha de la interacción del alumno con el material medido como el número de días desde el inicio del módulo de presentación	Atributo
Identificador Sitio			Un número de identificación del material.	Atributo
Tipo de actividad			El rol asociado con el material del módulo.	Atributo
Puntaje			El puntaje del estudiante en esta evaluación. El rango es de 0 a 100. La puntuación inferior a 40 se interpreta como Fallo. Las marcas están en el rango de 0 a 100.	Atributo
Intentos previos			El número de intentos que ha realizado el estudiante.	Atributo

h. Diagrama de relaciones binarias



i. Diccionario de clases

NOMBRE DEL CONCEPTO	INSTANCIAS	ATRIBUTOS DE CLASE	ATRIBUTOS DE INSTANCIA	RELACIONES
Estudiante			Nivel de educación, id estudiante, rango de edad, género, número de créditos	pertenece, tiene
Curso			identificador de curso, identificador de módulo	tiene
Interacción			cantidad interacción	
Evaluación por estudiante			identificador, puntaje	
País		nombre país		

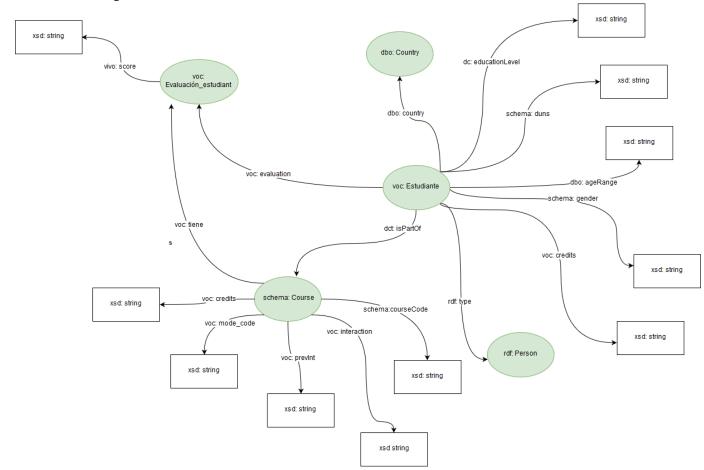
j. Matriz de atributos

NOMBRE DEL ATRIBUTO	CONCEPTO	TIPO DE VALOR	RANGO DE VALORES	CARDINALIDAD
Nivel de educación	Estudiante	Cadena de caracteres		1,1
Identificador Estudiante	Estudiante	Cadena de caracteres		1,1
Género	Estudiante	Cadena de caracteres		1,1
Código Modulo	Curso	Cadena de caracteres		1,1
Código Presentación	Curso	Cadena de caracteres		1,1
Cantidad Interacciones	Interacción	Número		1,1
Identificador Sitio	Entorno de aprendizaje	Cadena de caracteres		1,1
Tipo de actividad	Entorno de aprendizaje	Cadena de caracteres		1,1
Puntaje	Evaluación Estudiante	Número	1-100	1,1
Intentos previos	Evaluación Estudiante	Número	1-100	1,1

k. Tabla de Relaciones Binarias

NOMBRE DE LA RELACION	CONCEPTO ORIGEN DE DOMINIO	CARDINALIDAD MAXIMA	CONCEPTO DE DESTINO RANGO	RELACION INVERSA
Pertenece	Estudiante	N	País, Curso	
Tiene	Estudiante, Curso	N	Evaluación Estudiante, Interacción	

I. Modelo formal de la ontología

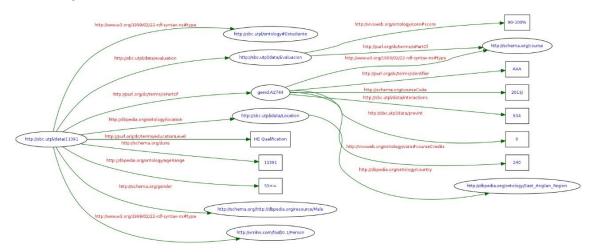


7. DESARROLLO

7.1. Análisis y Desarrollo de Ontología

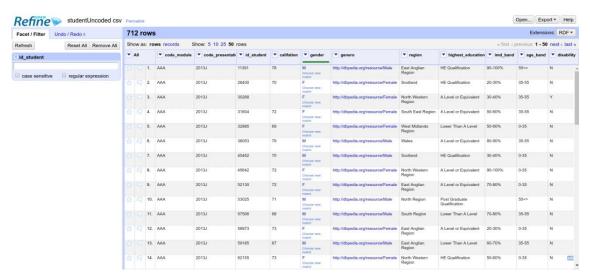
Para desarrollar la ontología se ha analizado un dataset establecido con el cual se ha definido las propiedades y clases, para crear la ontología.

La ontología ha sido desarrollada a través de Jena, la cual hemos obtenido como resultado.



7.2. Refinamiento y tratamiento de Datos

Para el tratado y refinamiento de datos se lo ha realizado a través de la herramienta Refine, el cual no has permitido darle asociar los datos con los recursos que están disponibles en la Dbpedia.



7.3. Publicación de la ontología en Virtuoso con Maquita Virtual en Digital Oceans

1. Ingreso por ssh a la máquina virtual.

```
→ ssh root@206.189.215.246
root@206.189.215.246's password: _
```

2. Instalación de Virtuoso 6

```
root@entornos-tesis:~# sudo apt-get install virtuoso-opensource
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following package was automatically installed and is no longer required:
  libodbc1
The following additional packages will be installed:
  virtuoso-vad-conductor virtuoso-vsp-startpage
Suggested packages:
  virtuoso-vad-bpel virtuoso-vad-dbpedia virtuoso-vad-demo virtuoso-vad-doc
  virtuoso-vad-facetedbrowser virtuoso-vad-isparql virtuoso-vad-ods virtuoso-vad-rdfmappers
  virtuoso-vad-sparqldemo virtuoso-vad-syncml virtuoso-vad-tutorial
The following NEW packages will be installed:
  virtuoso-opensource virtuoso-vad-conductor virtuoso-vsp-startpage
\theta upgraded, 3 newly installed, \theta to remove and 61 not upgraded. Need to get \theta B/1,606 kB of archives.
After this operation, 7,509 kB of additional disk space will be used. Do you want to continue? [Y/n]
```

3. Iniciar servicio de Virtuoso

```
root@entornos-tesis:~# sudo service virtuoso-opensource-6.1 start
root@entornos-tesis:-# sudo service virtuoso-opensource-6.1 status

• virtuoso-opensource-6.1.service - LSB: Virtuoso OpenSource Edition
Loaded: loaded (/etc/init.d/virtuoso-opensource-6.1; generated; vendor preset: enabled)
Active: active (running) since Sun 2019-03-17 23:36:45 UTC; 4s ago
Docs: man:systemd-sysv-generator(8)
Process: 15198 ExecStop=/etc/init.d/virtuoso-opensource-6.1 stop (code=exited, status=0/SUCCESS)
Process: 17041 ExecStart=/etc/init.d/virtuoso-opensource-6.1 start (code=exited, status=0/SUCCESS)
Tasks: 25 (limit: 4915)
CGroup: /system.slice/virtuoso-opensource-6.1.service
L7046 /usr/bin/virtuoso-t +wait +configfile /etc/virtuoso-opensource-6.1/virtuoso.ini

Mar 17 23:36:42 entornos-tesis systemd[1]: Starting LSB: Virtuoso OpenSource Edition...
Mar 17 23:36:45 entornos-tesis systemd[1]: Started LSB: Virtuoso OpenSource Edition.
lines 1-13/13 (END)
```

4. Instalar isql para conectarse a la base de datos de virtuoso.

```
root@entornos-tesis:~# sudo apt-get install unixodbc
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following NEW packages will be installed:
    unixodbc
0 upgraded, 1 newly installed, 0 to remove and 61 not upgraded.
Need to get 0 B/54.7 kB of archives.
After this operation, 111 kB of additional disk space will be used.
Selecting previously unselected package unixodbc.
(Reading database ... 48905 files and directories currently installed.)
Preparing to unpack .../unixodbc_2.3.4-1_amd64.deb ...
Unpacking unixodbc (2.3.4-1) ...
Setting up unixodbc (2.3.4-1) ...
```

5. Conectarse a la base de datos de virtuoso con usuario dba.

```
root@entornos-tesis:~# isql-vt -U dba

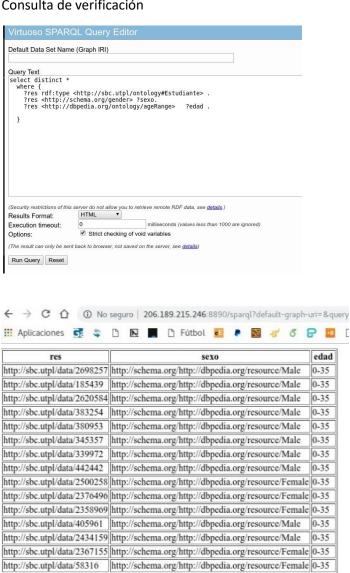
*** Error 28000: [Virtuoso Driver]CL034: Bad login
at line 0 of Top-Level:

Enter password for dba :
Connected to OpenLink Virtuoso
Driver: 06.01.3127 OpenLink Virtuoso ODBC Driver
OpenLink Interactive SQL (Virtuoso), version 0.9849b.
Type HELP; for help and EXIT; to exit.
SQL>_
```

6. Cargar el archivo rdf al EndPoint http://sbc-resources.com/learning

```
Oriver: 06.01.3127 OpenLink Virtuoso ODBC Driver
OpenLink Interactive SQL (Virtuoso), version 0.9849b.
Type HELP; for help and EXIT; to exit.
SQL> ld_dir('/tmp', '*.rdf', 'http://sbc-resources.com/learning');
Done. -- 2 msec
```

7. Consulta de verificación



7.4. Explotación y desarrollo de Aplicación para consulta de Datos

http://sbc.utpl/data/266524 http://schema.org/http://dbpedia.org/resource/Female 0-35 http://sbc.utpl/data/2293923 http://schema.org/http://dbpedia.org/resource/Male 0-35 http://sbc.utpl/data/447481 http://schema.org/http://dbpedia.org/resource/Male 0-35 http://sbc.utpl/data/58873 http://schema.org/http://dbpedia.org/resource/Female 0-35 http://sbc.utpl/data/102806 http://schema.org/http://dbpedia.org/resource/Male http://sbc.utpl/data/2494700 http://schema.org/http://dbpedia.org/resource/Male http://sbc.utpl/data/133922 http://schema.org/http://dbpedia.org/resource/Male 0-35 http://sbc.utpl/data/567922 http://schema.org/http://dbpedia.org/resource/Female 0-35 http://sbc.utpl/data/2165164 http://schema.org/http://dbpedia.org/resource/Female 0-35

Para el análisis de los datos desarrollados en RDF se ha utilizado el framework de desarrollo web Angular6, el cual permite realizar peticiones http al Endpoint de Virtuoso y analizar la ontología desarrollada previamente.

a) Función para realizar peticiones http a Virtuoso

b) Etapa donde se predefinen las consultas que permiten hacer la aplicación.

c) Parte de la interfaz en donde se escoge la consulta a realizar.

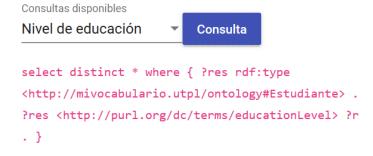
d) Resultado de la consulta ejecutada



8. CONSULTAS REALIZADAS

1. Nivel de Educación de los estudiantes.

Escoja una consulta

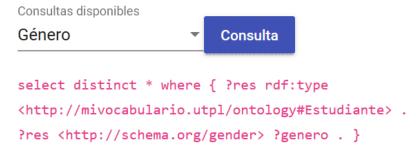


Resultado de la consulta ejecutada:

res	r
http://sbc.utpl/data/2698257	Lower Than A Level
http://sbc.utpl/data/2329342	Lower Than A Level
http://sbc.utpl/data/2613116	Lower Than A Level
http://sbc.utpl/data/2586560	Lower Than A Level
http://sbc.utpl/data/749412	Lower Than A Level
http://sbc.utpl/data/146503	Lower Than A Level
http://sbc.utpl/data/2012966	Lower Than A Level
http://sbc.utpl/data/2084066	Lower Than A Level
http://sbc.utpl/data/527801	Lower Than A Level
http://sbc.utpl/data/2456480	Lower Than A Level

2. Género de los estudiantes inscritos

Escoja una consulta

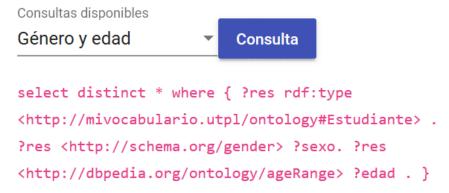


Resultado de la consulta ejecutada:

genero
nttp://schema.org/http://dbpedia.org/resource/Male

3. Conocer el género y rango de edad de los estudiantes.

Escoja una consulta



Resultado de la consulta ejecutada:

res	sexo	edad
http://sbc.utpl/data/2698257	http://schema.org/http://dbpedia.org/resource/Male	0-35
http://sbc.utpl/data/185439	http://schema.org/http://dbpedia.org/resource/Male	0-35
http://sbc.utpl/data/2620584	http://schema.org/http://dbpedia.org/resource/Male	0-35
http://sbc.utpl/data/383254	http://schema.org/http://dbpedia.org/resource/Male	0-35
http://sbc.utpl/data/380953	http://schema.org/http://dbpedia.org/resource/Male	0-35
http://sbc.utpl/data/345357	http://schema.org/http://dbpedia.org/resource/Male	0-35
http://sbc.utpl/data/339972	http://schema.org/http://dbpedia.org/resource/Male	0-35
		1

4. Saber el rango de calificaciones que han obtenido los estudiantes

Escoja una consulta

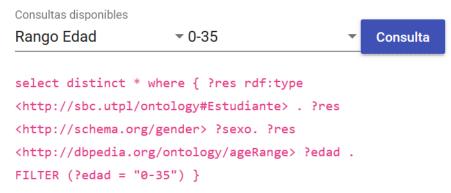


Resultado de la consulta ejecutada.

res	evaliacion	califacion
http://sbc.utpl/data/2698257	http://sbc.utpl/data/Evaluacion	0-10%
http://sbc.utpl/data/185439	http://sbc.utpl/data/Evaluacion	0-10%
http://sbc.utpl/data/130547	http://sbc.utpl/data/Evaluacion	0-10%
http://sbc.utpl/data/969076	http://sbc.utpl/data/Evaluacion	0-10%
http://sbc.utpl/data/2620584	http://sbc.utpl/data/Evaluacion	0-10%
http://sbc.utpl/data/409100	http://sbc.utpl/data/Evaluacion	0-10%
http://sbc.utpl/data/383254	http://sbc.utpl/data/Evaluacion	0-10%

5. Consultar los estudiantes por rango de edad

Escoja una consulta



Resultado de la consulta ejecutada.

res	sexo	edad
http://sbc.utpl/data/2698257	http://schema.org/http://dbpedia.org/resource/Male	0-35
http://sbc.utpl/data/185439	http://schema.org/http://dbpedia.org/resource/Male	0-35
http://sbc.utpl/data/2620584	http://schema.org/http://dbpedia.org/resource/Male	0-35
http://sbc.utpl/data/383254	http://schema.org/http://dbpedia.org/resource/Male	0-35
http://sbc.utpl/data/380953	http://schema.org/http://dbpedia.org/resource/Male	0-35

9. BIBLIOGRAFÍA

- 1. Fayyad, U.M.; Piatetsky-Shapiro, G.; Smyth, P.; Uthurusamy, Advances in Knowledge And Data Mining. (1996). MIT Press. Massachussets, USA
- 2. Lyn, Thomas; Edelman, David; Crook, Jonathan, Credit Scoring and its Applications. SIAM.Filadelfia, USA (2002).
- 3. Mladenic, D.; Grobelnik, M.; Jermol, M. Exploiting Text Mining in Publishing and Education. Workshop on Data Mining Lessons Learned. ICML. Pp. 34-39. 2002.
- 4. Minaei-Bidgoli, B.; Punch, W.F. Predicting student performance: an application of data mining methods with the educational web based system LON-CAPA. IEEE Frontiers in Education. Pp 1-6. 2003.
- 5. J. Senso, "Herramientas para trabajar con rdf." El profesional de la información 12.2 (2003): 132-139.
- 6. PLS RAMBOLL MANAGEMENT, Studies in the Context of the E-learning Initiative: Virtual Models of European Universities (Lot) (2004).
- 7. Jenkins M., Browne T., Walker R., "VLE Surveys. A longitudinal perspective between March 2001, March 2003 and March, 2005 for higher education in the United Kingdom" (2005).
- 8. R. C. Fernández, "Representación del Conocimiento. Web Semántica," Universidad Carlos III de Madrid, 2005.
- 9. Berners-Lee, T.; Hendler, J.; Lassila, O. "The semantic web: a new form of web content that is meaningful to computers will unleash a revolution of new possibilities". En: Scientific American, 2001, mayo.
- 10. Hendler, J.; Golbeck, J.; Parsia, B. Trust networks on the semantic web, 2002.
- 11. Biólogo, T. D. E., Ojeda, A. V., Alejandra, C., Chávez, B., Raimundo, Á., & Sc, M. (2015). La Universidad Católica de Loja.
- 12. Romero, C. y Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. Expert Systems with Applications, 33(1), 135-146. doi: 10.1016/j.eswa.2006.04.005
- 13. Bienkowski, M., Feng, M., & Means, B. (2012). Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief. Washington, D. C.: U.S. Department of Education.