Saber, Universidad de Oriente, Venezuela.Vol. 27 Nº 2: 293-299. (2015)
ISSN: 2343-6468 Digital / ISSN: 1315-0162 Impreso / Depósito Legal pp 198702SU187

CLUSTER INFORMÁTICO PARA MODELAR LA CORRELACIÓN DE LAS TIC Y LOS RESULTADOS ACADÉMICOS EN LA PRUEBA ENLACE

INFORMATIC CLUSTER FOR MODELING CORRELATION BETWEEN ICT AND ACADEMIC RESULTS OF THE ENLACE TEST

Verónica Quintero[1], Miguel Muñoz[2], Claudia Martínez[1], Arnoldo Díaz Ramírez[1]

[1]Instituto Tecnológico de Mexicali, [2]Centro de Bachillerato Tecnológico Industrial y de Servicios, Mexicali, Baja California, México. E-mail: veronicaquintero@itmexicali.edu.mx / adiaz@itmexicali.edu.mx

RESUMEN

En el marco de las nuevas tecnologías de información y comunicación (TIC) en la Educación Media Superior se realiza un estudio del impacto que estas nuevas tecnologías tienen respecto al rendimiento académico de los estudiantes. El estudio se realiza a través de un algoritmo de correlación y una aplicación informática que permite analizar y evaluar la relación de variables que existe entre el uso de estas herramientas y el desempeño obtenido en cada una de las competencias académicas. El aporte metodológico del presente estudio se inscribe en un análisis cuantitativo utilizando el método y algoritmo matemático de correlación lineal estadística a través de un cluster informático.

Palabras clave: Competencias, tecnologías, comunicación, información.

ABSTRACT

Within the framework of the new information and communication technologies (ICT) in the high school, this study is performed to evaluate the impact of these new technologies on the academic performance of students. The study was conducted through a correlation algorithm and a computer program to analyze and evaluate the relationship between variables using these tools and the performance obtained in each of the academic competences. The methodological contribution of this study is performed by a quantitative analysis using the method and mathematical algorithm of statistical linear correlation by means of an informatic cluster.

Key words: Competences, technology, communication, information.

INTRODUCCIÓN

En México anualmente se aplica una evaluación nacional de logro académico en centros escolares (ENLACE), es una prueba del sistema educativo nacional que se aplica a planteles públicos y privados del país. La prueba ENLACE se usa en educación básica (primaria y secundaria) y en Educación Media Superior. Esta evaluación cubre las competencias disciplinarias básicas de los campos de comunicación y matemáticas y cuyo propósito es generar una sola escala de carácter nacional que proporcione información comparable de los conocimientos y habilidades que tienen los estudiantes en los temas evaluados clasificándolos en cuatro diferentes tipos de nivel de dominio los cuales son **Insuficiente**, **Elemental**, **Bueno** y **Excelente.**

Anteriormente se han realizado algunos estudios de los resultados de la prueba ENLACE con respecto a la influencia que pudiesen tener otros parámetros como las calificaciones escolares independientes del sondeo ENLACE (Campos Vázquez y Urbina Romero 2011), el alcance de este tipo de estudios se limita al uso de datos confidenciales de estudiantes en diferentes regiones del país y solo a muestras proporcionadas por las mismas autoridades escolares sujetas al estudio estadístico que permitan hasta cierto grado determinar la relación existente entre ambos parámetros.

La propuesta de esta investigación se centró en la correlación que puede existir entre dos parámetros esenciales de la educación media superior, uno es el uso de las nuevas tecnologías de información y comunicación (TIC) y el segundo parámetro son los resultados académicos obtenidos por los estudiantes en la prueba ENLACE a nivel nacional. Los datos requeridos por cada uno de los parámetros son de libre acceso, el uso e indicadores de las TIC se obtienen del Instituto Nacional de Geografía Estadística e Informática de México (INEGI 2014) y los resultados académicos a nivel nacional se obtienen de ENLACE (SEPM 2014).

Aunque se puede tener acceso libre a los datos de los parámetros planteados, la relación de los datos es muy compleja y se requiere de un método manual para sorteo y filtrado de información ya sea para un análisis global o

Recibido: noviembre 2014. Aprobado: febrero 2015.
Versión final: abril 2015.

regional y esto conlleva a una gran cantidad de manejo de datos que pueden generar errores muy significativos. Para dar solución a esta problemática, en esta investigación se ha propuesto un cluster informático para modelar la correlación existente entre los parámetros antes mencionados utilizando un algoritmo de correlación de datos en combinación con algoritmos de calendarización (Berman *et al*. 2003), para el manejo de gran cantidad de información así como para la optimización de tiempo maquina con el uso de procesamiento paralelo.

VARIABLES DE LA PRUEBA ENLACE Y LAS TIC

ENLACE clasifica los resultados obtenidos por nivel de dominio**,** cada uno de los indicadores se clasifica en cuatro diferentes niveles los cuales son **Insuficiente**, **Elemental**, **Bueno** y **Excelente**, evaluando así mismo las habilidades de comunicación y matemáticas; considérese entonces los cuatro niveles de dominio anteriores como las variables $\boldsymbol{Yi_1}$, $\boldsymbol{Yi_2}$, $\boldsymbol{Yi_3}$, $\boldsymbol{Yi_4}$, respectivamente.

En lo que respecta al uso de las TIC, los indicadores más significativos de acuerdo a ICT (Doriska 2009), son el uso de la **computadora** para la sección de Tecnologías de Información, el uso del teléfono **celular** para la sección de Tecnología de Telecomunicaciones y el uso del **Internet** para la sección de Tecnología de Redes; considérese entonces estos tres indicadores como las variables $\boldsymbol{Xi_1}$, $\boldsymbol{Xi_2}$, $\boldsymbol{Xi_3}$, respectivamente.

Para realizar el análisis de la relación entre las variables, se requiere conocer con que fuerza influye una variable con el comportamiento de la segunda variable, es decir, que tan importante es una variable y que tanto influye en el resultado de una correlación estadística; en términos generales la correlación estadística determina la relación o dependencia que existe entre las dos variables que intervienen en una distribución bidimensional. La relación de variables planteadas anteriormente se observa en la Tabla 1.

Tabla 1. Relación de variables TIC y ENLACE.

TIC	*Variables Xi*	**ENLACE**	*Variables Yi*
Xi_1	Computadora	Yi_1	Insuficiente
Xi_2	Celular	Yi_2	Elemental
Xi_3	Internet	Yi_3	Bueno
		Yi_4	Excelente

ALGORITMO DE CORRELACIÓN

En una distribución bidimensional puede ocurrir que las dos variables guarden algún tipo de relación entre sí, el coeficiente correlación mide el grado de intensidad de esta posible relación entre las variables cuantitativas y se calcula aplicando la siguiente fórmula (Pita Fernández y Pértega Díaz 1997):

$$r = \frac{\frac{1}{n}\Sigma\left(Xi_j - \overline{X}\right)\left(Yi_k - \overline{Y}\right)}{\left(\left(\frac{1}{n}\Sigma\left(Xi_j - \overline{X}\right)^2\right)_{j=1,2,3}\left(\frac{1}{n}\Sigma\left(Yi_k - \overline{Y}\right)^2\right)\right)_{k=1,2,3,4}^{\frac{1}{2}}} \quad (1)$$

Donde,

- *n es el número de muestras a correlacionar*
- *Xi, Yi son números reales positivos*
- *j, k definen el tipo de indicador a correlacionar*

Los valores que puede tomar el coeficiente de correlación "r" son: -1 < r < 1

- Si "r" > 0, la correlación lineal es positiva (si sube el valor de una variable sube el de la otra). La correlación es tanto más fuerte cuanto más se aproxime a 1.

- Si "r" < 0, la correlación lineal es negativa (si sube el valor de una variable disminuye el de la otra). La correlación negativa es tanto más fuerte cuanto más se aproxime a -1.

- Si "r" = 0, no existe correlación lineal entre las variables.

Los valores de *Xi* son porcentajes de uso de las TIC en México y los valores de *Yi* son porcentajes obtenidos de los alumnos ubicados por nivel de dominio según resultados obtenidos en la prueba ENLACE en cada una de las entidades federativas del país. La Figura 1 muestra un ejemplo de datos estadísticos factibles para un análisis de grado de correlación en el área de Matemáticas teniendo como referente el uso de Internet (Xi_3) en relación a los niveles de dominio de ENLACE: Insuficiente, elemental, bueno, excelente ($\boldsymbol{Yi_1}$, $\boldsymbol{Yi_2}$, $\boldsymbol{Yi_3}$, $\boldsymbol{Yi_4}$).

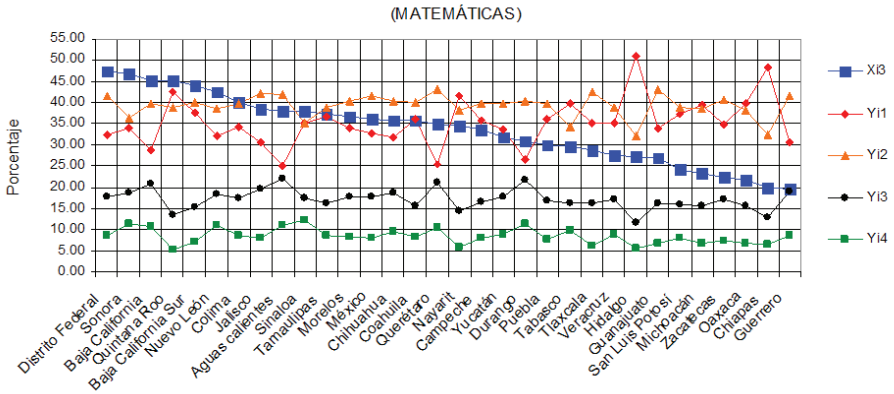

Figura 1. Datos estadísticos para las variables de las TIC y ENLACE (Fuente: INEGI 2012).

REPARTICIÓN DE TAREAS EN EL CLUSTER

El método para modelar la correlación de las TIC y la prueba ENLACE en Bachillerato es a través de un algoritmo de calendarización utilizando un clúster de tres computadoras configuradas con el sistema operativo Red Hat de Linux. Para la distribución de tareas se utilizó el método *Messaging Program Intarce* (MPI)/*Parallel Virtual Machine* (PVM). Es importante destacar que se pueden configurar una mayor cantidad de nodos (Casanova 2001), el servidor principal lo reconocerá y le asignará tareas, la configuración propuesta se muestra en la Figura 2.

Previamente el algoritmo se ejecutó en un procesador normal (computadora HP Pavilion, procesador 3.4 GHz, 4 GB SDRAM, 1024 GB 7200 rpm de disco duro). Este algoritmo, se volvió muy pesado, con una latencia excesiva, y un tiempo de respuesta muy lento, esto se hizo analizando la respuesta de master slave dependiendo de la respuesta que dio en cada corrida, es decir, llamaremos muestra 1 cuando el proceso solo arroje una tarea terminada, muestra 2 cuando el proceso arroje dos tares terminadas y así sucesivamente. La respuesta del modelo de correlación fue lenta pero después de la iteración (iteración = milisegundo) cuarenta y uno, empezó a tener una latencia muy grande, como se muestra en la Figura 3.

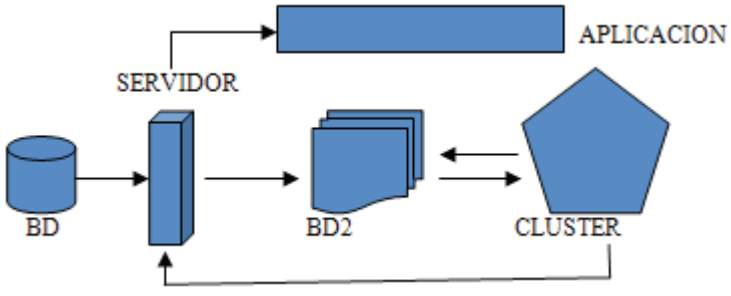

Figura 2. Arquitectura del algoritmo de calendarización

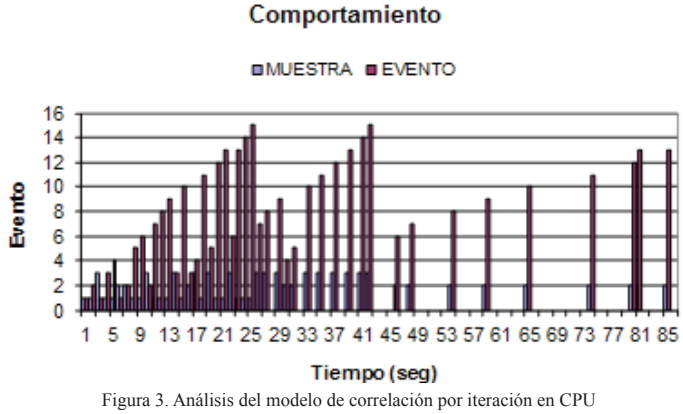

Figura 3. Análisis del modelo de correlación por iteración en CPU

El siguiente paso fue analizar el comportamiento de este mismo algoritmo utilizando el clúster propuesto. La planificación consistió entonces en el despliegue de las tareas de un trabajo sobre nodos del sistema, en nuestro caso fueron únicamente dos nodos de tal forma que el rendimiento final dependerá de los siguientes factores:

- Concurrencia: uso del mayor número de procesadores simultáneamente.
- Grado de paralelismo: el grado más fino en el que se pueda descomponer la tarea.
- Costes de comunicación: diferentes entre procesadores dentro del mismo nodo y procesadores en diferentes nodos.
- Recursos compartidos: uso de recursos comunes para varios procesadores dentro del mismo nodo.

En cuanto a las reglas para la administración y balanceo de carga en el Cluster:

- Administrar la disponibilidad de los nodos.
- Configurar atributos de los nodos que sean importantes para el balanceo de cargas.
- Configurar y diseñar políticas.
- Administrar reservaciones y recursos dedicados.
- Monitorear y registrar un historial de uso de recursos para usuarios y grupos.

En el presente caso, no se incluyeron credenciales de autentificación o seguridad ya que es interno, sin embargo esto debe de implementarse si se requiere que el cluster tome datos de internet, como lo es base de datos del INEGI, en este caso el candado y seguridad lo tendría el nodo principal es decir el servidor. El anexo A muestra la metodología utilizada de la asignación de tareas en el cluster.

ANEXO A (Asignación de tareas en cluster informático)

cexec
Ejecuta un comando en todos los nodos. # cexec "ps | grep ps.txt"

cget
Copia archivos de una cierta ubicación en los nodos. Ignora enlaces y directorios. Si existe un nombre de archivo con el mismo nombre lo renombra con un sufijo formado por el nombre del nodo del *cluster*.

cget /etc/rc.d/rc.local

ckill
Permite finalizar un proceso en ejecución en los nodos del *cluster*. Para utilizarlo se utiliza el nombre del proceso, y no su ID de proceso debido que en cada nodo el ID es diferente.

ckill –u talkd log.txt

cpush
Permite mover archivos de una cierta ubicación en los nodos. # cget /home/local /home/rc.bk

crm
Permite eliminar archivos y directorios en los nodos. Su funcionamiento es similar al comando **rm**, con las opciones de interactivo y recursivo.

crm –iR /home/dafa/ver1/

cshutdown
Permite apagar, reiniciar o suspender un nodo. Las opciones son las mismas del comando shutdown en un sistema Linux. Adiciona el uso de la opción t para especificar el tiempo que tomará para ejecutar la acción.

cshutdown r t 0

clist , cget , cnum

Son utilizados para hacer consultas de archivos de configuración

Se utilizó una planificación estática y no dinámica debido a que no se definió el equilibrio de cargas ya que son solo dos nodos y no existe el equilibrio de conexiones ni migración de datos.

La definición de la toma de archivos en la base de datos para la correlación esta dada por:

```
chost# clzonecluster show -v sczone
chost# clresource delete -F -Z sczone hasp-rs
chost# clzonecluster configure sczone
czone> remove dataset name=HAzpool
czone> commit
chost# clzonecluster show -v sczone
```

Finalmente los resultados en el modelado de la correlación en el cluster fueron más estables con respecto al método anterior ya que el tiempo de respuesta fue casi inmediato a los procesos requeridos, en este caso no existieron tiempos muertos o fueron casi nulos como se muestra en la Figura 4.

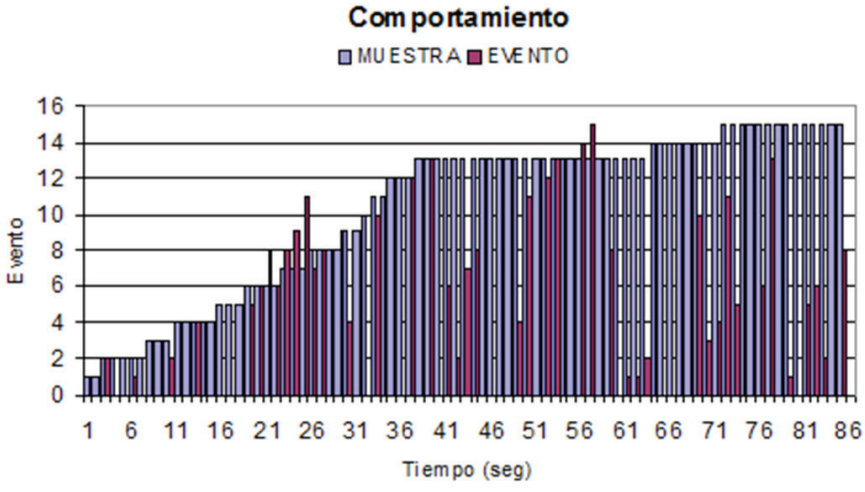

Figura 4. Análisis del modelo de correlación ejecutado en cluster informático.

RESULTADOS

La ejecución de los algoritmos con el uso del cluster informático arrojaron datos interesantes en el modelo de correlación estadística representado por coeficientes. Los coeficientes fueron el resultado de la relación entre las dos variables planteadas en la sección dos y el grado o fuerza con la que influye una variable con respecto a la otra se denota por el valor numérico obtenido por el coeficiente. Las figura 5, 6 y 7 muestran los coeficientes de correlación obtenidos para cada uno de los indicadores de las TIC con relación en las variables de la prueba ENLACE.

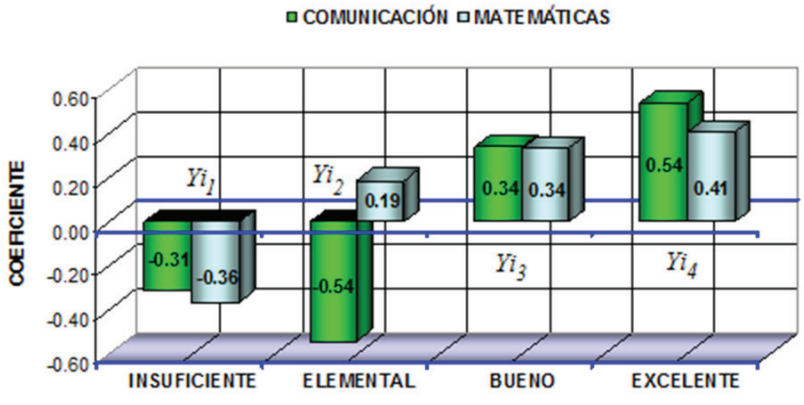

Figura 5. Coeficientes de correlación entre la variable Xi_1(computadora) y variables de la prueba ENLACE.

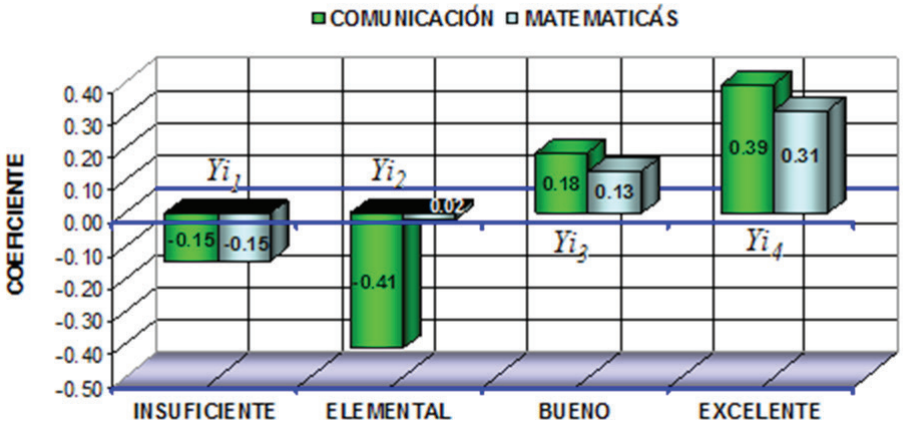

Figura 6. Coeficientes de correlación entre la variable XI_2 (celular) y variables de la prueba ENLACE.

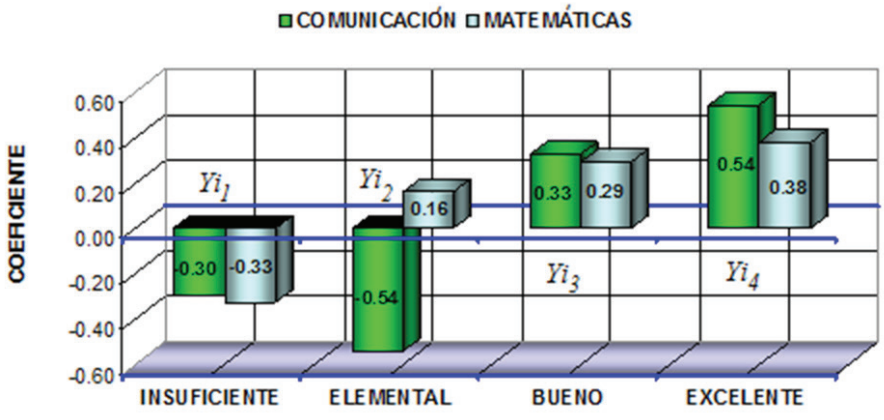

Figura 7. Coeficientes de correlación entre la variable Xi_3 (Internet) y variables de la prueba ENLACE.

CONCLUSIONES

Se ha propuesto un cluster informático para modelar la correlación de los resultados académicos de ENLACE en Educación Media Superior y con relación al porcentaje de uso de las TIC de los estudiantes. Cabe mencionar que los algoritmos planteados en esta investigación no están limitados al uso exclusivo de los resultados de ENLACE, cualquier otro indicador de resultados académicos tales como (PISA, EXCALE, entre otros) son factibles para modelarse en el cluster informático. Con relación a los datos académicos, el modelo proporciona datos de correlación estadística que pueden auxiliar como indicadores confiables según el grado de correlación obtenido en cada variable para la distribución y uso eficiente de las nuevas tecnologías de información y comunicación en las distintas regiones del país acorde al impacto y necesidades arrojados por los resultados académicos.

La principal característica por la cual fue utilizado el modelo informático se debió a que, a través del procesamiento paralelo, el resultado se define mediante el cálculo simultáneo de una ecuación con múltiples datos a procesar. De manera que el tiempo para resolver un problema se reduce a la asignación de las tareas a los recursos de cómputo con lo cual asegura que se satisfagan y a todas las relaciones de precedencia en las tareas (ecuación). Así mismo el cluster informático deja abierta la posibilidad para ser utilizada con aplicaciones móviles las cuales requieren de tiempos de procesamiento menores para ser funcionales.

REFERENCIAS BIBLIOGRÁFICAS

CAMPOS VÁZQUEZ R, URBINA ROMERO F. 2011. Desempeño educativo en México: La prueba enlace. Estudios Económicos. 26(2):249-292.

CASANOVA H. 2001. SimGrid: for simulation of application A toolkit scheduling. IEEE/ACM International Symposium on Cluster Computing and the

Grid. Dept. of Comp. Sci. & Eng., California Univ., San Diego, La Jolla, CA. DOI: 10.1109/CCGRID.2001.923223

Doriska MW. 2009. Partnership on Measuring for Development. Revisions and Additions to the Core List of ICT Indicators. Statistic Commission Background document.

Pita Fernández S, Pértega Díaz S. 1997. Relación entre variables cuantitativas. Cad. Aten. Primaria. 4:141-144.

Berman F, Rey G, Fox C, Hey T. 2003. The Grid: past, present, future. *In*: Berman F, Fox G, Hey T. (Ed). Grid Computing: Making the Global Infrastructure a Reality. John Wiley & Sons Inc, Chichester, USA, pp. 1060.

INEGI (Instituto Nacional de Geografía Estadística e Informática de México). 2012. Gobierno de México. Disponible en línea en: http://www.inegi.org.mx (Acceso 03.03.2014).

INEGI (Instituto Nacional de Geografía Estadística e Informática de México). 2014. Gobierno de México. Disponible en línea en: http://www.inegi.org.mx (Acceso 03.03.2014).

SEPM (Secretaria de Educación Pública de México). 2014. ENLACE: Evaluación Nacional para el Logro Académico. Disponible en línea en: http://www.elance.sep.gob.mx (Acceso 05.03. 2014).