

Sistema experto para calificar pruebas de desarrollo en estudiantes de la Universidad Continental

Expert system to evaluate development tests in the Universidad Continental students

Rolando Antonio Párraga Chamorro¹
Universidad Continental

RESUMEN

Objetivos: Construir un sistema experto para calificar pruebas de desarrollo en Informática, y comprobar si su aplicación permite la calificación de una prueba de desarrollo con una valoración cuantitativa que permita al alumno autoevaluarse en cualquier momento recibiendo información instantánea sobre su nivel de conocimientos.

Métodos: El tipo de investigación fue aplicada, de campo, descriptiva y transversal. EXAINFO tiene básicamente dos partes: la base de conocimientos y el motor de inferencia para obtener una valoración del examen de desarrollo. Para la evaluación de su aplicabilidad se construyó la matriz de evaluación del curso, luego el modelo de características para extraer la puntuación y las palabras vacías, de este modo se calificó respuestas construidas de los estudiantes e indicó los conceptos faltantes en el aprendizaje; en este proceso se cuidó el principio de concurrencia, mediante la validación de captura facial, evitando así la suplantación de algún estudiante. La muestra fue de 40 estudiantes del primer semestre de la asignatura de Informática de la Universidad Continental. **Resultados:** El promedio de 72,38 % de tasa de recuperación indica que la mayor parte de los conceptos básicos son recuperados por los estudiantes. Los conceptos básicos recuperados son demasiados por lo que se reduce la tasa de precisión. **Conclusiones:** El sistema experto EXAINFO sí hace posible calificar las pruebas de desarrollo y diagnosticar el estado de aprendizaje del alumno mediante el análisis de texto libre de las respuestas del estudiante.



Rolando Párraga

rparraga@continental.edu.pe

Historial del artículo:

Recibido: 16 de abril de 2014

Aprobado: 14 de noviembre de 2014

Disponible en línea: 30 de diciembre de 2014

Palabras clave: Sistema experto, calificar, pruebas de desarrollo, informática.

¹ Magíster en Ingeniería de Sistemas, docente e investigador de la Universidad Continental.

ABSTRACT

Objectives: To build an expert system to evaluate development tests in informatics and to prove if its application enables the development tests evaluation with a quantitative score which allows students to evaluate themselves at any time and receive instant information about his/her knowledge level. **Methods:** The research type was applied, field research, descriptive and transversal. Exainfo basically has two parts: the knowledge base and the inference engine to obtain a test development score. For the applicability evaluation, we built the subject evaluation matrix, after the features model to obtain the rating and empty words, in that way we qualify students built responses and indicate the missing concepts in learning; in this process we were careful with the concurrence principle by facial capture validation, to prevent student impersonation. The sample was formed by 40 Universidad Continental first semester students from the Informatics subject. **Results:** The recovery rate average of 72,38 % indicates that most of the basic concepts are recovered by students. The retrieved basic concepts are too many so the accuracy rate is reduced. **Conclusions:** The Exainfo expert system makes possible to qualify development tests and diagnose the student learning state through the free text analysis of student responses.

Keywords: Expert system, qualify, development test, informatics.

INTRODUCCIÓN

Los sistemas expertos constituyen programas que reproducen el proceso intelectual de un experto humano en un campo particular. En teoría pueden razonar siguiendo los pasos que seguiría un experto humano para resolver un problema concreto. La utilización de los sistemas expertos para el aprendizaje pueden ser eficaces, ya que es de esperar que el usuario pueda aprender observando el comportamiento del sistema (1).

En la prueba objetiva, los estudiantes

reciben una puntuación o calificaciones para representar su estado de aprendizaje. La información no es adecuada para los estudiantes ni para los profesores, ya que no pueden identificar explícitamente los conceptos que faltan de los estudiantes. Lo que implica proporcionar a los estudiantes los conceptos insuficientes de aprendizaje después de la prueba.

La evaluación de las respuestas de los estudiantes en las pruebas de desarrollo es una actividad que requiere mucho tiempo para los profesores. En el proceso enseñanza-aprendizaje, la evaluación del aprendizaje juega un papel importante, ya que determina la medida en que se están cumpliendo los objetivos del proceso.

Como indicador del aprendizaje, se asume el nivel de conocimientos y habilidades que el alumno posee en el área temática de interés (materia, tema, etc.). Los educadores utilizan diferentes medios para medir el nivel del aprendizaje de los temas que enseñan. Las evaluaciones propias de nuestras universidades comprenden, mayoritariamente, exámenes escritos (sean parciales o finales). El alumno al final de un determinado semestre académico recibe la medición "aprueba" o "desaprueba" la asignatura, pero sin ser identificados explícitamente los conceptos que a ellos les faltó aprender en la respectiva asignatura.

Los exámenes computarizados para la calificación de las respuestas construidas del estudiante en función de las preguntas abiertas, representan un gran avance en la aplicación de la evaluación formativa, a través de diagnósticos de aprendizaje entregados a los estudiantes de una manera oportuna que les permita dirigir sus procesos de aprendizaje y puedan planear sus rutas de estudio de conceptos al invocar un examen en línea de respuesta construida.

Backhoff e Ibarra (2) presentan la primera etapa de una línea de investigación de las universidades de muchos países industrializados donde los exámenes computarizados empiezan a sustituir a las evaluaciones tradicionalmente diseñadas

para lápiz y papel. Describen la versión computarizada del examen de habilidades y conocimientos básicos.

Bunderson, Inouye y Olsen (3) manifiestan que la nueva tecnología digital empieza a diseñar y desarrollar nuevos modelos más sofisticados de evaluación. Ellos hablan de cuatro generaciones de la evaluación asistida por computadora y que cada una de estas representa un avance sobre la otra, lo que implica un incremento en su poder y sofisticación:

- **Primera generación:** evaluación computarizada (test informatizados).
- **Segunda generación:** evaluación adaptativa (test adaptativos informatizados).
- **Tercera generación:** evaluación continua (sistemas tutoriales, estima los cambios en la trayectoria de aprendizaje curricular del estudiante).
- **Cuarta generación:** evaluación inteligente (sistemas tutoriales inteligentes, produce, interpreta y genera perfiles de los resultados del estudiante con base en conocimiento y procedimientos de inferencia).

Hwang (4) propuso un modelo basado en mapas concepto efecto, llamado CER, que demostraría que el aprendizaje de ciertos conceptos se ve influenciado por el aprendizaje de otros conceptos previos. De esta manera, el modelo entregaba al estudiante un diagnóstico que le guiaba a través de los conceptos que debían ser profundizados. El modelo recién generado se probó en un curso de Ciencias Naturales de nivel básico primario en la República de Taiwán.

En el 2009, Hwang, Panajabure, Triampo y Shin (5) proponen un algoritmo para facilitar la asignación de pesos (o incidencia numérica) a cada uno de los conceptos involucrados en el mapa de concepto efecto de su modelo CER. Dan cuenta de que es uno de los pasos más dispendiosos y de mayor incidencia en el diagnóstico que se le

entrega al estudiante, por lo que se diseña un algoritmo que permite, de una manera metódica, asignarle a cada concepto involucrado en un proceso de aprendizaje, un peso numérico que signifique el grado de importancia que cada concepto tiene para un aprendizaje específico por parte del alumno.

En general, en su gran mayoría existe software para calificar exámenes de prueba objetiva que compara directamente las respuestas de los estudiantes con las respuestas estándar. Desafortunadamente, este método está restringido a tener que preparar respuestas estándar con antelación y es difícil asegurar que el estudiante carece de información de aprendizaje. Para corregir este problema, contamos con las características que se utilizan en el método de valorar las respuestas de texto libre y comparar con la matriz de aprendizaje en forma automática para ayudar a la tarea de calificar una prueba de desarrollo automáticamente en tiempo real.

Por esta razón y frente a la necesidad de dar alternativas de calificar una prueba de desarrollo y de diagnosticar el aprendizaje a la situación descrita, surgió el siguiente planteamiento: ¿Es posible construir un Sistema Experto para calificar pruebas de desarrollo en Informática en los estudiantes de la Universidad Continental?

El objetivo general del estudio fue diseñar un sistema experto para la calificación de pruebas de desarrollo basadas en la extracción de las características de un determinado tema para comparar con las respuestas construidas del estudiante y generar información sobre el nivel de conocimientos que cada uno de ellos posee en la asignatura de Informática.

La hipótesis general sostiene que sí es posible construir un sistema experto para calificar pruebas de desarrollo en Informática y su aplicación permite obtener una medición de aprobado o desaprobado recibiendo información instantánea sobre el nivel de los conocimientos que poseen.

MATERIAL Y MÉTODOS

La investigación fue de tipo básica y aplicada, pues se ha diseñado un software y luego se ha aplicado para comprobar su utilidad.

Según Hernández Sampieri (6), la investigación tiene un diseño no experimental.

Para la aplicación del software Exainfo se cuidó el principio de concurrencia al laboratorio de cómputo, mediante el instrumento biométrico de captura facial a fin de evitar la suplantación del alumno.

El estudio identificó patrones de relación entre las variables a partir de la matriz de evaluación, integrado a la base de datos SQL derivados del software Exainfo para calificar la prueba de desarrollo y diagnosticar el aprendizaje del estudiante con una prueba no estandarizado de reactivos de respuesta construida.

La población estuvo constituida por todos los alumnos del primer semestre de las carreras profesionales que ofrece la Universidad Continental. La muestra fue de 40 estudiantes procedentes de la asignatura de Informática I, la que representó 40 pruebas de desarrollo.

Entre las variables, la independiente fue: Software Exainfo para calificar pruebas de desarrollo; la dependiente: Aprendizaje del estudiante; y las intervinientes estuvieron constituidas por el nivel de reactivos de respuesta construida (medio y alto).

La técnica empleada fue la recolección de reactivos de los exámenes de pruebas de desarrollo aplicados a los estudiantes en semestres anteriores. Otra técnica fue la observación del comportamiento del estudiante durante la prueba de desarrollo en el laboratorio de cómputo.

Como instrumentos estuvo la batería de instrumentos aplicados a los estudiantes de la asignatura de Informática I, constituida por una prueba de desarrollo para evaluar

el nivel de aprendizaje de un determinado tema, y una lista de hechos principales que evalúa el grado de dificultad de los reactivos que permite inferir las respuestas del examinado. Es de aplicación individual y posee validez de contenido.

Los equipos utilizados fueron el terminal biométrico ZKTeco para captura facial y huellas, PC, laptops; y las técnicas de procesamiento de datos, el TF, TF-IDF, entropía, core-precision y el core-recall.

Elaboración de la base de conocimiento

Exainfo es un sistema específico para evaluación y está estructurado por materias o asignaturas independientes, cada una de las cuales tiene una base de conocimientos distinta. La base de conocimientos de Exainfo está formada por tres tipos de objetos: 1) los conceptos, que son los ítems o elementos en los que se descompone la asignatura en contenidos que se muestra en la tabla N° 1, y están estructurados jerárquicamente, de acuerdo con un sílabo basado en objetivos instruccionales/competencias; 2) los tests, que representan las sesiones de evaluación y que cada uno de ellos está formado por un conjunto de ítems o cuestiones; y 3) los ítems o cuestiones, que están asociados a uno o varios conceptos o temas. La definición de los tests se hace en función de los temas sobre los que se desea evaluar, todas ellas calibradas con una serie de parámetros.

La búsqueda de conceptos que representan el conocimiento del experto está basada en la información que aporta el experto según la matriz de evaluación de la asignatura de Informática I, como muestra la tabla N° 2.

Los reactivos de respuesta construida se almacenó en una base de datos SQL-Server a través de LINQ TO SQL, realizado así por su conocimiento profundo del lenguaje Visual C# en el cual se desarrolló el software Exainfo.

Para la etapa de construcción del motor de inferencia del concepto efecto, contamos con la colaboración de docentes de la

Tabla N° 1: Tabla de contenido de Informática I.

Contenido	Conocimientos
I. Fundamentos de informática	1. Fundamentos de informática 2. Recursos y componentes del sistema 3. Suites ofimáticas 4. Hardware y software 5. Sistema operativo 6. Servicios de Internet
II. Procesador de texto	7. Documentos sencillos 8. Documentos avanzados y plantillas
III. Software de presentaciones	9. Diseñador de presentación 10. Diapositivas y plantillas
IV. Software de diagramación	11. Formas 1D y 2D 12. Diagramas y plantillas
V. Hoja electrónica de cálculo	13. Libros sencillos 14. Funciones 15. Datos 16. Macros

Facultad de Ingeniería.

Los patrones de respuesta construida fueron elaborados siguiendo algunas recomendaciones consignadas en el documento "A new approach for constructing the concept map", de cinco investigadores taiwaneses de las universidades National ChiaoTung University y Asia University (7).

Al generar un test representa una sesión de evaluación en función del tema o temas a evaluar. Los ítems correspondientes a un test serán los necesarios para realizar la evaluación. No existe una asociación directa entre test e ítem, salvo a través de los temas. Se impone además que cada ítem está asociado no solo al tema para el que fue definido, sino a todos los ascendentes dentro de la jerarquía de temas.

El generador de test es el encargado de seleccionar las preguntas a plantear al alumno, según el entendimiento del aprendiz sobre los conceptos expuestos en la tabla 2, para incluir las carencias y malas concepciones de aprendizaje mostradas en la estructura de la base de conocimientos (figura N° 1).

Rendir una prueba

Para hacer una prueba solo hay que seguir los siguientes pasos:

- Validar la captura facial mediante acceso biométrico facial
- Iniciar sesión en el sistema Exainfo
- Pulsar el botón "rendir prueba"
- Completar la prueba
- Ver resultado de la prueba

Calificación de la prueba de desarrollo

Exclusión de puntuación y palabras vacías

Para la recolección de la información útil a partir de las respuestas de los estudiantes, el primer paso del proceso consiste en eliminar la puntuación, números y palabras vacías, ya que sirven como ruido de los roles.

Stemming

Stemming es un procedimiento de reducción de las palabras a sus elementos mínimos con significado de una palabra flexionada (o derivado en algún momento) de base o

Tabla N° 2: Matriz de evaluación de Informática I (período 2013-II).

Semana	Contenidos	Habilidades	Nivel de dificultad		
			Medio	Alto	Puntaje
1	Fundamentos de la informática. Recursos y componentes del computador. Identificación de recursos y componentes del sistema. Hardware y software.	Explica los conceptos básicos de la informática. Identifica los elementos de un computador.	X		2
2	Sistemas operativos. Instalación y configuración de aplicaciones informáticas. Aplicaciones ofimáticas y corporativas. Suites ofimáticas. Licencias.	Compara y reconoce los diferentes tipos de software. Utiliza las herramientas de administración del sistema operativo.		X	2
3	Internet. Servicios de Internet. Correo electrónico. Fuente de recursos. Tendencias actuales	Reconoce la conexión de una red. Explica los servicios de Internet.	X	X	4 4
4	Procesador de textos. Documentos sencillos. Formatos. Inserción de objetos. Herramientas de revisión.	Elabora documentos sencillos. Inserta diversos objetos a un documento.	X		4
5	Procesador de textos. Documentos avanzados. Plantillas. Combinar correspondencia. Estilos. Tabla de contenidos. Macros.	Realiza la combinación de correspondencia Inserta tablas de contenidos	X		4
6	Aplicaciones ofimáticas de presentación. Iniciar una presentación nueva. Trabajar texto de la diapositiva. Ajustar el diseño, orden y apariencia de las diapositivas.	Crea presentaciones para una exposición. Aplica efectos de transición y progresión de diapositivas	X		4
7	Fundamentos de herramientas de diagramación. Entorno de dibujo. Formas 1D y 2D.	Identifica las distintas herramientas de diagramación. Utiliza formas 1D y 2D.	X		2
8	Diagramas y plantillas.	Combina el uso básico de herramientas y las plantillas para la creación de diagramas.		X	3
9	Examen parcial.				
10	Hoja de cálculo. Documentos sencillos: manejar hojas de cálculo. Dar formato a las celdas.	Realiza operaciones básicas.	X		3
11	Funciones: fecha, texto, matemáticas y estadísticas. Gráficos estadísticos.	Aplica las funciones de fecha, texto, matemáticas y estadísticas.	X		4
12	Funciones lógicas. Funciones condicionales.	Aplica las funciones lógicas y condicionales.	X		4
13	Funciones de búsqueda y referencia.	Aplica las funciones de búsqueda y referencia.	X		3
14	Datos. Validación de datos. Formato condicional. Filtros. Subtotales.	Resume grandes cantidades de datos.	X		2
15	Agrupar y esquematizar datos. Tablas y gráficos dinámicos. Importar y exportar datos.	Inserta la tabla dinámica para resumir la información.		X	3
16	Hoja de cálculo avanzado. Plantillas.	Conoce los diferentes tipos de plantillas.	X		4
17	Macros en libro electrónico de cálculo.	Utiliza macros para automatizar procedimientos repetitivos.		X	5

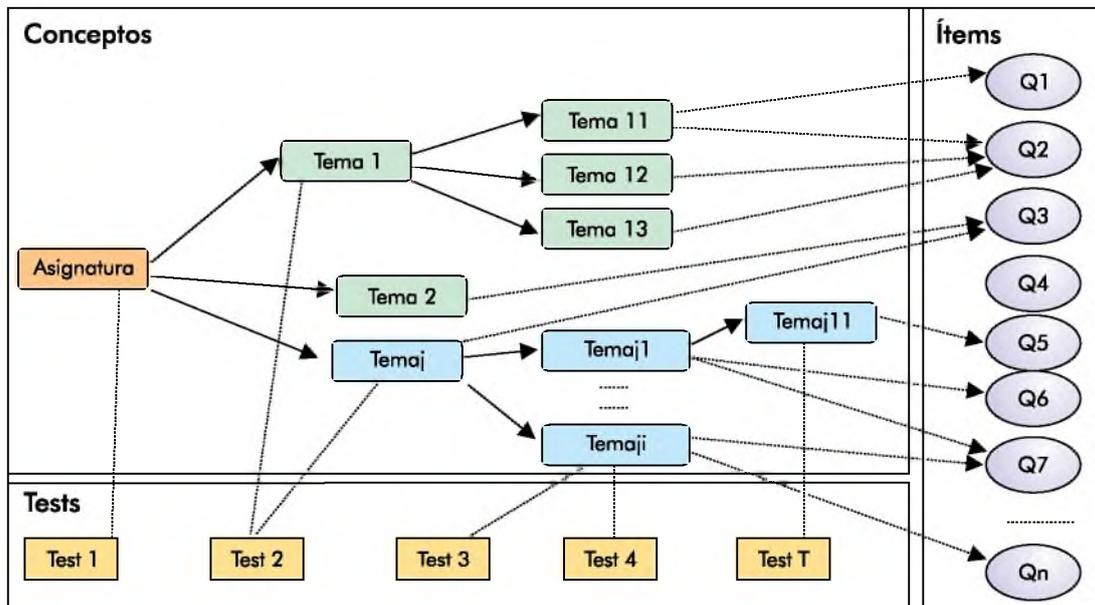


Figura N° 1: Estructura de la base de conocimientos.

forma de la raíz. Por lo general, la palabra derivada no tiene que ser idéntica a la raíz morfológica de la palabra, según el algoritmo de Porter (8).

Extracción de características

Las características (en este caso una palabra) extraídas de las respuestas son los términos frecuentes (TF), la inversa del término frecuente y el documento de la frecuencia (TF-IDF), y la entropía-variación (EV), que es la cantidad de información de cada palabra entre la respuesta del estudiante.

La frecuencia de la palabra en la respuesta dada es simplemente el número de veces que aparece un término dado en esa respuesta. Para calcular la frecuencia de la palabra en particular en la respuesta α_i del término t_i , se calcula el término frecuencia $ff_{i,j}$ de la siguiente manera (9):

$$ff_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,i}} \quad (1)$$

Donde:

$n_{i,j}$ es el número de veces que aparece el término t_i en la respuesta α_j , y el producto

es la suma del número de ocurrencias de todos los términos en la respuesta α_j .

El peso TF-IDF es una medida estadística que se utiliza con frecuencia en la recuperación de información por medio de la minería de datos (extracción de información y representación de los mismos para obtener los patrones).

La fórmula de cálculo de la frecuencia inversa de documento idf_i , para el término t_i es el siguiente:

$$idf_i = \log \frac{|D|}{|\{d:t_i \in D\}|} \quad (2)$$

Donde:

$|D|$ es el número de documentos en la colección, y $|\{d:t_i \in D\}|$ es el número de documentos donde aparece el término t_i . Si el término no está en la colección, se producirá una división-por-cero. Por lo tanto, es común ajustar esta fórmula a $(1 + |\{d:t_i \in D\}|)$.

Matemáticamente, la base de la función logaritmo no es importante y constituye un factor constante en el resultado final. Luego, TF-IDF se calcula como:

$$(tf-idf)_{i,i} = tf_{i,i} \times idf_i \quad (3)$$

Donde $tf_{i,i}$ y idf_i son definidos en ecuaciones (1) y (2), respectivamente.

La entropía es la incertidumbre media de una única variable aleatoria. La definición de la entropía $H(x)$ es expresada en términos de un conjunto discreto de probabilidades $p(x_i)$:

$$H(x) = -\sum_{i=1}^n p(x_i) \log_2 p(x_i) \quad (4)$$

En este trabajo se utiliza el concepto de entropía, y queremos cuantificar la cantidad de información de cada palabra entre las respuestas de los estudiantes. Lo llamamos como entropía-variación (EV). En nuestro estudio, x_i se asigna a la palabra x en la respuesta del alumno i th, y $p(x_i)$ es la probabilidad de que la palabra x en la respuesta del alumno i th, debido a que hay n respuestas de ellos, i es de 1 a n . Por lo tanto, $H(x)$ representa la incertidumbre media de la palabra x aparición en las respuestas. Para evitar el valor EV que tiende a tener grandes valores como el número de alumnos incrementados, que normalizan $H(x)$ de modo que el valor está comprendido entre 0 y 1. La fórmula de la $H(x)$ normalizada, llamado $normH(x)$, se indica a continuación:

$$normH(x) = \frac{1}{n} \sum_{i=1}^n p(x_i) \log_2 p(x_i) \quad (5)$$

Separación de datos

En el estudio experimentamos con la validación cruzada de 5 veces, la muestra original es dividida aleatoriamente en cinco submuestras. De estas submuestras, una única submuestra se mantiene como validación de datos para las pruebas, y las cuatro submuestras restantes se utilizan como datos de entrenamiento. La validación cruzada en el proceso se repite cinco veces. Los cinco resultados se promedian para producir una única estimación.

Característica procedimiento de comparación

En esta fase, queremos encontrar conceptos que son importantes en las respuestas, primero comparamos las características del TF, TF-IDF y EV entre los datos de entrenamiento y los datos de prueba. Si el valor de la característica es mayor que cero en los datos de entrenamiento, pero es cero en los datos de prueba, entonces vamos a extraer esa característica como la falta candidato al concepto central. Después de ello clasificamos los candidatos al concepto central que faltan según los valores en los datos de entrenamiento. Por último, porque nuestro corpus contiene las preguntas de definición donde las respuestas no son largas, seleccionamos en la mayoría de los diez conceptos fundamentales que faltan con los valores de características superiores.

Experimentación

Para aplicar el sistema Exainfo, se involucró 17 conceptos y aplicó 40 pruebas de desarrollo en Informática I, con seis preguntas de texto libre para cada estudiante.

RESULTADOS

Las métricas de evaluación que adoptamos son el ratio de precisión y el ratio de recuperación. Ambos tipos son ampliamente utilizados para evaluar los sistemas en el dominio del procesamiento de lenguaje natural. El ratio de precisión es el número de conceptos recuperados pertinentes dividido por el número total de conceptos recuperados en una búsqueda de base de datos. El ratio de recuperación es la proporción relevante a todos los conceptos pertinentes en la base de datos.

Preparamos y analizamos los conceptos de respuesta libre y hacemos algunas modificaciones, por dos razones. En primer lugar, el número de conceptos fundamentales en una definición no es muy grande, por lo que recuperamos los

primeros cinco o diez conceptos que faltan. En segundo lugar, si el estudiante obtiene una puntuación más alta en la prueba, significa que los conceptos que faltan son los no aprendidos, así que tenemos que incluir el número de los conceptos centrales correctas que el estudiante posee durante la evaluación. Nuestros indicadores de evaluación son llamados core-precision y core-recall. Core-precision@5 y core-precision@10 representan cinco y diez conceptos propuestos respectivamente. Las fórmulas son las siguientes:

$$\text{core_precision@5} = \frac{\text{match} + \text{correct}}{\text{match} + \text{correct}} \quad (6)$$

$$\text{core_precision@10} = \frac{\text{match} + \text{correct}}{10 + \text{correct}} \quad (7)$$

$$\text{core_precision} =$$

$$\frac{\text{core_precision@5} + \text{core_precision@10}}{2} \quad (8)$$

$$\text{core_recall} = \frac{\text{match} + \text{correct}}{\#\text{Keyword}} \quad (9)$$

Donde:

Match es el número de los conceptos propuestos por nuestro sistema, que son idénticos a los conceptos básicos patrones dados por el docente.

Correct es el número de conceptos básicos que el alumno tiene en la respuesta.

#Keyword es el número total de los conceptos básicos que el profesor da.

Experimentamos con seis preguntas de texto libre para cada estudiante en relación con las respuestas correctas sobre el número de preguntas disponibles en la prueba, los resultados finales se muestran en las tablas N° 3 y tabla N° 4.

Tabla N° 3: Resultados experimentales core-precision.

Pregunta/ Característica	Preg. 1	Preg. 2	Preg. 3	Preg. 4	Preg. 5	Preg. 6	Promedio (%)
TF	53,58 %	36,73 %	10,13 %	15,45 %	53,03 %	56,57 %	37,58 %
TFIDF	52,09 %	43,95 %	10,13 %	15,45 %	52,56 %	57,18 %	38,56 %
EV	52,58 %	26,74 %	4,37 %	15,45 %	50,93 %	49,97 %	33,34 %
TF+TFIDF	52,09 %	42,27 %	10,48 %	15,45 %	52,56 %	57,53 %	38,40 %
TF+EV	52,58 %	27,69 %	4,73 %	15,45 %	50,93 %	49,62 %	33,50 %
TFIDF+EV	52,58 %	26,74 %	4,37 %	15,45 %	50,93 %	49,62 %	33,28 %
ALL	52,58 %	27,69 %	4,37 %	15,45 %	50,93 %	49,62 %	33,44 %

Tabla N° 4: Resultados experimentales core-recall.

Pregunta/ Característica	Preg. 1	Preg. 2	Preg. 3	Preg. 4	Preg. 5	Preg. 6	Promedio (%)
TF	89,30 %	61,62 %	71,95 %	56,36 %	76,74 %	70,15 %	71,02 %
TFIDF	87,51 %	70,48 %	69,55 %	56,36 %	76,34 %	70,55 %	71,80 %
EV	89,00 %	46,99 %	28,79 %	56,36 %	74,74 %	64,16 %	60,01 %
TF+TFIDF	87,51 %	68,81 %	74,32 %	56,36 %	76,34 %	70,95 %	72,38 %
TF+EV	89,00 %	47,95 %	33,57 %	56,36 %	74,74 %	63,95 %	60,93 %
TFIDF+EV	89,00 %	46,99 %	31,16 %	56,36 %	74,74 %	63,95 %	60,37 %
ALL	89,00 %	47,95 %	31,16 %	56,36 %	74,74 %	63,95 %	60,53 %

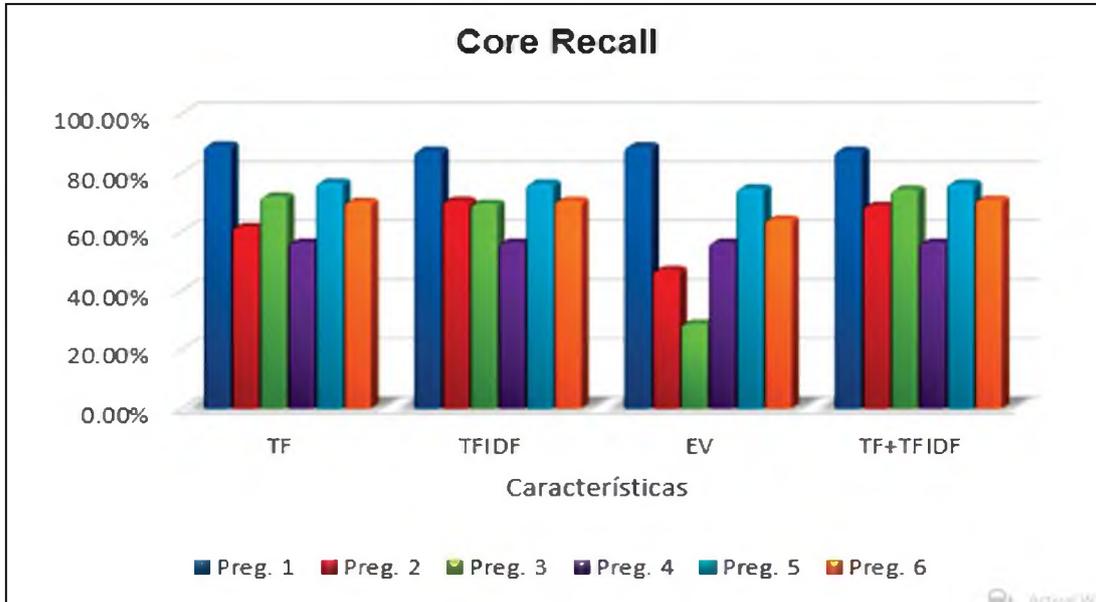


Figura N° 2: Resultados experimentales según el core-recall.

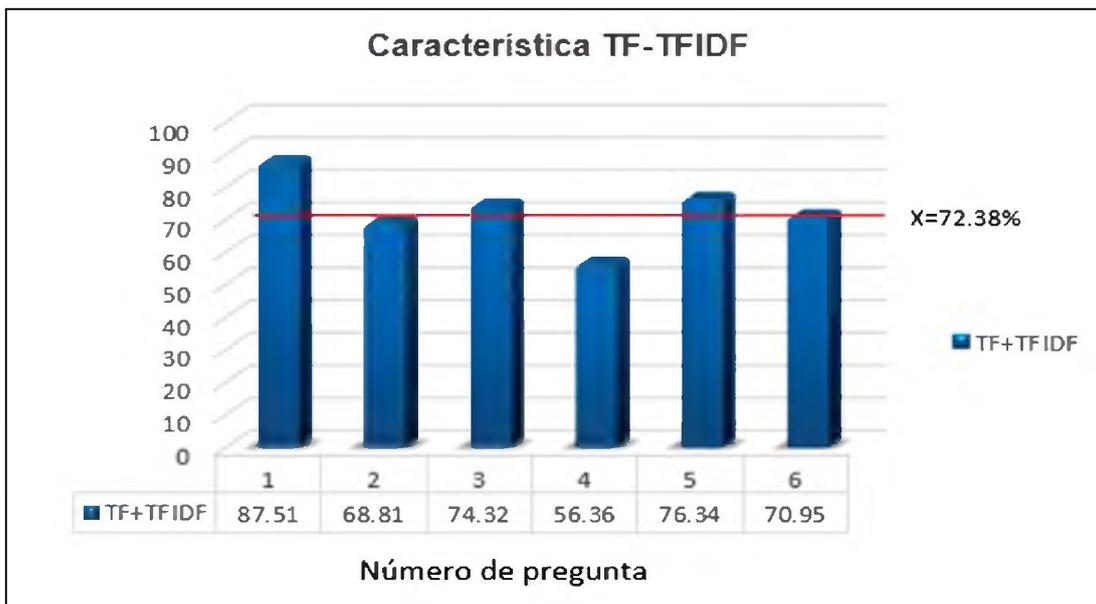


Figura N° 3: El core-recall usando la frecuencia de los términos TF-TFIDF como característica.

En el experimento, primero tratamos las características individuales y luego hacemos combinaciones entre ellas. En las características individuales, el TFIDF realiza el mejor performance tanto en el core-precisión y como en el core-recall (figura 3). En las diferentes combinaciones de las características, en cambio, el TF+TFIDF alcanza el mejor performance en el core-recall, con un ratio promedio de 72,38 % (figura N° 3), lo que significa que la

mayor parte de los conceptos faltantes son recuperados por los estudiantes.

Analizamos el resultado en la figura N° 3, y podemos descubrir que las actuaciones para las preguntas 2, 4 y 6 no alcanzan el ratio promedio de core-recall. La razón para que este resultado se dé en el caso de la pregunta 2, según el estudio es que la mayoría de las respuestas en esta pregunta fueron con puntuaciones inferiores a

trece, en las respuestas, los estudiantes solo escribieron el esbozo de solución sin explicación detallada de manera que las respuestas fueron similares a la respuesta estándar en la matriz de referencia. El resultado en la pregunta 4 fue el más bajo porque la respuesta estándar era muy corta y los alumnos confundieron con la respuesta correcta (función contar por contará). En el caso de la pregunta 6 ocurrió porque era un problema más difícil de Middleware, la mayoría de los estudiantes obtuvieron calificaciones inferiores a 12. Se provocó que los ejemplos de hardware fueran más que los de software.

Como resultado de las respuestas de texto libre consideramos que los conceptos correctos en la respuesta son pocos y que los alumnos responden demasiados conceptos básicos, por lo que se reduce el ratio de precisión.

Las respuestas se valoran, se almacenan y se transmite por medios electrónicos libres de error, para su análisis estadístico y la toma de decisiones sobre el aprendizaje del estudiante.

DISCUSIÓN

Los resultados expuestos evidencian que la mayoría de la muestra de la población estudiantil posee puntuaciones inferiores a 13, obtiene calificación en tiempo real con el software Exainfo y muestra los conceptos que faltan desarrollar en su aprendizaje. En consecuencia, debemos advertir que hay una diferencia entre calificar pruebas objetivas y calificar pruebas de desarrollo valorando las respuestas. En tal sentido, es ilustrativo el trabajo de Backhoff e Ibarra (2), que aplicó hace 20 años la primera versión computarizada de examen de habilidades y conocimiento básico que sustituye a las evaluaciones tradicionalmente diseñadas para lápiz y papel.

Exainfo es un software que captura las imágenes faciales de los estudiantes matriculados en la asignatura y así permite el acceso único al examen. Este

procedimiento toma como base el trabajo de investigación de nuevos modelos sofisticados de evaluación desarrollados por Bunderson, Inouye y Olsen (3).

El acceso al examen desarrollado por captura facial, tiene una serie de ventajas, como evitar la suplantación sobre todo por alumnos de ciclos superiores, y contabilizar la cantidad de exámenes que rindieron.

Nuestros hallazgos respecto a nivel de conocimientos se contradicen con los de Hwang (4), que demostraría que el aprendizaje de ciertos conceptos se ve influenciado por el aprendizaje de otros conceptos previos. Por la naturaleza de la asignatura, que es eminentemente práctica, cada aplicación no está vinculada con otra.

Al examinar las pruebas de desarrollo, se han encontrado entre otros aspectos que la valoración de las respuestas construidas está en función de las características extraídas del término de frecuencia (TF), para compararlos con los datos de entrenamiento; sin embargo, se contradice con la propuesta de Hwang, Panajabure, Triampo y Shin (5), que facilita la asignación de pesos (o incidencia numérica) para respuestas largas a cada uno de los conceptos involucrados en el mapa de concepto efecto de su modelo CER. Por consiguiente, en la práctica encontramos que las respuestas de los estudiantes a las preguntas abiertas son cortas, solucionando la valoración según el peso numérico asignado a la pregunta y considerando el grado de importancia que cada concepto tiene para un aprendizaje específico por parte del estudiante.

Los resultados experimentales del corecall indica el número de preguntas que respondieron incorrectamente y alerta los conceptos faltantes en la enseñanza-aprendizaje, que serán nuestros indicadores para diseñar nuevas estrategias de enseñanza o programar una retroalimentación de las clases impartidas.

En conclusión, el software Exainfo sí hace posible calificar las pruebas de desarrollo

en Informática en los estudiantes de la Universidad Continental, validados por captura facial, debido a que extrae características de las respuestas de texto libre del estudiante y hace un análisis con las respuestas referenciadas del docente, y como tal los valores devueltos son datos en bruto. Considerando que el conocimiento de ciertos conceptos es causa de las respuestas a los ítems y del conocimiento de otros temas o conceptos relacionados, es poco probable que un alumno que ha demostrado conocimiento sobre un tema no tenga conocimiento sobre otro que se considera un prerrequisito de este. Por lo tanto, el TF y TFIDF son buenos indicadores para el diagnóstico de aprendizaje del Exainfo. Los maestros saben que se debería hacer, pero no siempre lo hacen; los niveles y conceptos erróneos de aprendizaje de los estudiantes son fuentes principales para realizar una retroalimentación de la enseñanza-aprendizaje.

Agradecimientos

A la Universidad Continental por el financiamiento brindado a la investigación; a la Facultad de Ingeniería por las facilidades para el desarrollo del software, y a los catedráticos, personal administrativo y a los jóvenes que con diversas tareas contribuyeron a la realización de este estudio.

REFERENCIAS BIBLIOGRÁFICAS

1. Durkin J. *Expert System Design and Development*. New Jersey: Prentice-Hall; 1994.
2. Backhoff E, Ibarra M. Versión computarizada del examen de habilidades y conocimientos básicos. 23 Congreso Internacional de Psicología Aplicada. Madrid; 1994.
3. Bunderson C, Inouye D, Olsen J. *The Four Generations of Computerized Educational Measurement*. 3a ed. New York: MacMillan Publishing; 1993.
4. Hwang G.J. A conceptual map model for developing intelligent tutoring systems. *Computers & Education*. 2003; 40: 217-235.
5. Hwang G.J, Panjaburee P, Triampo W, Shih B.Y. A multi-expert approach for developing testing and diagnostic systems based on the concept effect model. *Computers & Education*. 2009; 55: 527-540.
6. Hernández R, Fernández C. *Metodología de la Investigación*. 2a ed. México: McGraw-Hill; 2000.
7. Shian-Shyong T, Pei-Chi S, Jun-Ming S, Jui-Feng W, Wen-Nung T. A new approach for constructing the concept map. *Computers & Education*. 2005; 45(4): 9-14.
8. Porter M.F. An algorithm for suffix stripping. *Program*. 2006; 40(3): 211-218.
9. Salton G, McGill M.J. *Introduction to Modern Information Retrieval*. New York: McGraw-Hill; 1983.
10. Papineni K, Roukos S, Ward T, Zhu W. J. BLEU: A method for automatic evaluation of machine translation. New York: IBM Research Center; 2001.