

# Fundamentos e Implementación de un Sistema de Corrección Automatizada de Exámenes Utilizando Grafos Dirigidos

María Alejandra Paz Menvielle (pazmalejandra@gmail.com), Mario Alberto Groppo (proyale@groppo.com.ar), Marcelo Martín Marciszack (marciszack@gmail.com), Analía Guzmán (analia.guzman@the-group.com.ar), Karina Ligorria (karinaligorria@gmail.com), Martín Casatti (mcasatti@gmail.com)

*CIDS-Centro de Investigación, Transferencia y Desarrollo de Sistemas de Información  
Departamento de Ingeniería en Sistemas de Información Facultad Regional Córdoba –  
Universidad Tecnológica Nacional*

*Maestro Marcelo López esq. Cruz Roja Argentina Córdoba 0351 4686385*

**Resumen**—Se presenta el prototipo de un sistema de evaluación automática, que permite calificar exámenes formados por preguntas que se responderán con redacción de texto libre, en un dominio acotado, por alumnos de nivel universitario.

El prototipo implementa como mecanismo de corrección la búsqueda de patrones en una base de datos de grafos, la que se utiliza como base de conocimiento para la corrección automatizada de exámenes escritos por estudiantes en formato de texto.

Como parte del trabajo, se presentan los resultados preliminares obtenidos de su aplicación analizando los algoritmos propuestos para la determinación de la exactitud de las respuestas a través de la realización de exámenes de muestra, corregidos y calificados por docentes de la Cátedra Paradigmas de Programación y contrastados con los resultados que arroja el prototipo del sistema.

**Palabras clave**—análisis de texto; grafos; evaluación automática; detección de patrones

## 1. CONTEXTO

EL presente trabajo forma parte del proyecto de investigación y desarrollo homologado por la Secretaría de Investigación, Desarrollo y Posgrado de la Universidad Tecnológica Nacional, desarrollado en el ámbito del CIDS – Centro de Investigación, Desarrollo y Transferencia en Sistemas de Información, dentro del Departamento de Ingeniería en Sistemas de Información de la Facultad Regional Córdoba de la Universidad Tecnológica Nacional.

El dominio de aplicación seleccionado para la validación de la presente propuesta, se corresponde con los contenidos mínimos fijados para la asignatura Paradigmas de Programación[1], pertenecen al bloque de tecnologías básicas dentro del área programación y están principalmente referidos a los paradigmas lógicos, funcional y de orientación a objetos.

Este trabajo surge como parte de la evolución del equipo de investigación y desarrollo, según reflejan los trabajos

presentados en ediciones previas de CoNaIISI [2, 3, 4] expone los resultados obtenidos a través de una evaluación realizada por medio del prototipo desarrollado.

El objetivo del presente trabajo es describir brevemente el prototipo y presentar los resultados preliminares obtenidos de su aplicación, a través de la realización de exámenes modelos corregidos por docentes de la Cátedra Paradigmas de Programación y contrastados con los resultados que arroja el prototipo del sistema.

## 2. INTRODUCCIÓN

La evaluación, en sentido general se entiende por, “aquel conjunto de procesos sistemáticos de recogida, análisis e interpretación de información válida y fiable, que en comparación con una referencia o criterio nos permita llegar a una decisión que favorezca la mejora del objeto evaluado” [5, 6].

De este concepto se resaltan tres aspectos [7]. En primer lugar, evaluar es un proceso que se basa en una metodología y en técnicas. De aquí surge el concepto de medición. Sin medición, la sola valoración induce a una opinión subjetiva, no se estaría realizando una evaluación objetiva.

En segundo lugar, se evalúa justamente cuando se está en condiciones de establecer una comparación entre la información de que se dispone y alguno de los marcos de referencia o criterios que se establecieron para la evaluación.

Y, por último, el proceso de evaluación concluye con una toma de decisiones [8, 9]. En el concepto actual de evaluación, este es uno de los aspectos que está teniendo cada vez más importancia, sobre todo porque se intenta enlazar el proceso de evaluación con el proceso de mejora y, por tanto, la toma de decisiones se debería realizar con la finalidad de optimizar el proceso de evaluación.

La evaluación es una parte imprescindible dentro del proceso enseñanza-aprendizaje, como medida de la consecución de los objetivos de aprendizaje por parte del alumno,

y, también, como control de la calidad de dicho proceso. Sin embargo, los instrumentos de evaluación actuales presentan muchas limitaciones en el contexto lecto-escritor, que pueden ser superadas, en parte, por las nuevas tecnologías de la información y la comunicación[7].

Respecto al uso de la informática en el proceso de evaluación de aprendizajes, existen trabajos presentados sobre el uso de las tecnologías de información y comunicaciones (TIC) en el ámbito universitario que ponen de manifiesto que si bien la implementación de dichas tecnologías provee resultados indiscutiblemente beneficiosos[10] también es cierto que se encuentra una notable resistencia a la utilización de las TIC, tanto de parte del profesorado como de los alumnos.

El desafío consiste, entonces, en brindar un beneficio tangible a docentes o alumnos, que compense la carga de trabajo adicional y la resistencia inicial a la implementación de un nuevo método o tecnología.

En "Mirando el futuro: evolución de las tendencias tecnopedagógicas en educación superior"[11], Gros y Noguera mencionan que los métodos de evaluación "deben cambiar al mismo tiempo que los métodos de enseñanza, las herramientas y los materiales", o corren el riesgo de quedar relegados y resultar ineficientes en la evaluación de conocimientos obtenidos por medio de nuevos métodos y técnicas pedagógicas.

En este sentido, se presentarán los resultados obtenidos mediante la aplicación de una metodología, implementada en un prototipo automatizado, que permite calificar exámenes formados por preguntas que se responderán con redacción de texto libre, en un dominio acotado, por parte de alumnos de nivel universitario, el cual utiliza herramientas tecnológicas pero a la vez mantiene la capacidad expresiva y conceptual del lenguaje escrito coloquial.

Para lograr dicha interpretación expresiva y conceptual se aplicó el procesamiento de patrones, que es un ámbito de crucial importancia en todos los tópicos referidos a automatización de procesos. Tanto desde el punto de vista de detección de patrones desconocidos como desde el reconocimiento de patrones previamente detectados. Concretamente para este caso, se realizó el estudio sobre la representación del conocimiento utilizando grafos conceptuales.

Un grafo conceptual[12] es un sistema de notación simbólica y de representación del conocimiento. Presentado por John F. Sowa, se basa en los gráficos existenciales[13] de Charles Sanders Peirce, en las estructuras de redes semánticas y en datos de la lingüística, la filosofía y la psicología[14].

El enfoque que se eligió para realizar la búsqueda de patrones dentro de la información almacenada en forma de grafo, es el de búsquedas inexactas para la determinación de la validez, total o parcial, de las respuestas obtenidas.

En este enfoque se debe especificar al algoritmo, además de la estructura que se desea encontrar, un cierto umbral que debe ser tenido en cuenta si la estructura exacta no se encuentra. Teniendo en cuenta este umbral, el algoritmo busca los elementos de la estructura especificados con cierto grado de semejanza, por ejemplo con conceptos faltantes, con relaciones similares pero no idénticas, o con órdenes alterados o invertidos en las relaciones[15].

El mencionado enfoque se eligió debido a la naturaleza inherentemente variable de las respuestas registradas en

exámenes al mismo conjunto de preguntas.

Al momento de establecer esas relaciones conceptuales y, seguidamente, compararlas para obtener una evaluación de una respuesta dada, es necesario que todos los conceptos y relaciones involucrados se encuentren en el mismo espacio de dominio. Es decir, que todos esos conceptos y relaciones se hallen representados en el mismo grafo.

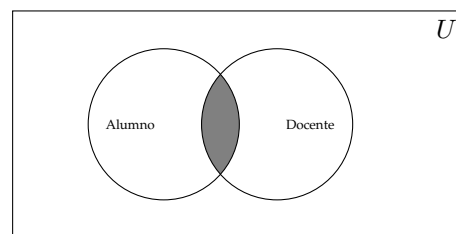


Figura 1: Respuesta con coincidencia débil

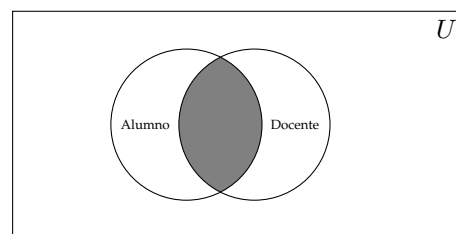


Figura 2: Respuesta con coincidencia fuerte

La evaluación surge de comparar la exactitud de la intersección (ver Figura 1 y 2) que se encuentra entre los conceptos incluidos por el docente en su respuesta ideal (denominada *respuesta base*) y los conceptos suministrados por el alumno en su respuesta propuesta (denominada *respuesta candidata*). En este contexto, una *coincidencia fuerte* tendrá una calificación superior a una *coincidencia débil*.

Actualmente se ha desarrollado un prototipo de sistema informático que implementa como mecanismo de corrección la búsqueda de patrones en una base de datos de grafos, la que se utiliza como base de conocimiento para la corrección automatizada de exámenes escritos por estudiantes en formato de texto.

Este mecanismo representa, de manera adecuada y simple, las estructuras del lenguaje natural con el fin de determinar si la respuesta a una pregunta es correcta dentro de un dominio especificado y aplicando las técnicas y las reglas gramaticales relacionadas con los lenguajes estructurados[16].

### 3. MÓDULOS DEL PROTOTIPO

El prototipo del sistema se compone de cinco módulos que son los bloques fundamentales que interactúan para obtener el resultado esperado, es decir, la evaluación de un examen redactado en texto libre.

#### 3.1. Módulo Gestión de Textos

Analiza las oraciones ingresadas al sistema buscando detectar y corregir errores ortográficos y sintácticos. También realiza análisis sintáctico de las palabras. Por la naturaleza

de las tareas que incluye, este módulo ha sido dividido en dos submódulos:

- **Ortográfico:** Analiza cada palabra ingresada al sistema con el fin de corregir su ortografía, en caso de no encontrar un par en la base de datos sugiere palabras relacionadas que el usuario puede seleccionar si la que considera correcta. La salida de este módulo es una oración sin errores ortográficos, ni léxicos, ni gramaticales.
- **Simplificador:** Desglosa la oración para almacenarla en la base de datos. En este proceso se asocia cada palabra con sus sinónimos o palabras equivalentes. Identifica en la oración corregida aquellas palabras fuertes para el dominio, esas palabras son principalmente sustantivos, adjetivos o palabras reservadas. También reconoce los conectores o relaciones que son relevantes en el significado de la respuesta que se busca.

### 3.2. Módulo Gestión de Conceptos

Administra la base de conocimiento, representada como un grafo dirigido, además permite obtener subconjuntos de conceptos y relaciones que los unen, que modelen respuestas tanto del docente como del alumno, permite la visualización y representación gráfica de los conceptos y relaciones y permite el proceso de evaluación de las respuestas para ver el grado de correlación entre la respuesta del docente y la del alumno. Dividido en los siguientes submódulos:

- **Administración de la base de conocimiento (Concept Manager):** Permite dar de alta, baja, modificación y consulta de conceptos y relaciones en el grafo.
- **Gestión de respuestas:** Permite el ingreso y la validación de una respuesta. Esta infraestructura sirve para la respuesta del docente y el alumno. Además si es una respuesta del docente (respuesta modelo) se realiza la grabación en el grafo, si es una respuesta del alumno se devuelve la representación de la respuesta en el grafo, es decir, la ruta que une los conceptos que representan la respuesta del alumno.
- **Graficador:** Permite la visualización y representación gráfica de los conceptos y relaciones existentes en el grafo.
- **Evaluador:** Determina que tan próxima es la respuesta del alumno con la respuesta modelo del docente, obtiene el peso o puntaje de la respuesta que servirá para determinar que tan exacta es la misma en relación con la respuesta modelo del docente.

### 3.3. Módulo Gestión de Exámenes

Permite administrar un examen, entre las operaciones básicas están: agregar, modificar, eliminar y consultar el conjunto de preguntas, establecer las respuestas correctas a las mismas y asignar los puntajes de cada una.

### 3.4. Módulo Académico

Permite gestionar los elementos académicos intervinientes en un examen, por ejemplo: docentes, alumnos, programas de estudio, unidades temáticas, etc.

### 3.5. Módulo Analítico

Permite realizar el análisis de los datos, por ejemplo, los conceptos más utilizados, los conceptos menos utilizados, las rutas típicas de respuestas, los errores más comunes, la cantidad de aciertos del sistema comparado con la forma de resolver del docente, etc. Este módulo será de fundamental importancia durante la etapa de prueba e implementación del sistema.

## 4. OPERATORIA DEL PROTOTIPO

Antes de describir la operatoria es necesario identificar los usuarios del sistema, es decir, aquellos que lo utilizarán con el fin de realizar el proceso de evaluación, en este caso los docentes y los alumnos.

Para realizar la automatización del proceso de evaluación se han identificado las responsabilidades que les corresponden a docentes y alumnos, para poder utilizar la herramienta automática:

#### Docente

- Elaboración de un banco de preguntas en función de criterios de evaluación previamente definidos.
- Elaboración y asignación de las respuestas relacionadas a las preguntas del banco de preguntas (preguntas base).
- Especificación de los criterios de corrección (equivalencias, ponderación, etc.) de cada respuesta.
- Selección de preguntas para componer una prueba de evaluación.
- Especificación de los criterios de corrección (equivalencias, ponderación, etc.) de cada respuesta.
- Validación y confirmación de la corrección automática realizada por el sistema. El docente puede aceptar una respuesta como válida, aún cuando el sistema la clasifique como dudosa, tal como se detalla en la sección de evaluación de respuestas.

#### Alumno

- Selecciona el examen correspondiente.
- Responde a las preguntas que se les presenten en el examen correspondiente.
- Consulta sus calificaciones y el estado de sus respuestas.

La operatoria general del prototipo del sistema posee un conjunto de pasos básicos para lograr la corrección automatizada de exámenes.

En la operatoria a describir se utilizarán dos conceptos que son de uso general en todos los módulos involucrados y que sirven para explicar la operatoria general:

- **Respuesta base:** Respuesta suministrada por el docente a una pregunta de examen.
- **Respuesta candidata:** Respuesta suministrada por el alumno en relación a una pregunta de examen.

Cabe aclarar que el sistema no requiere que el alumno brinde una respuesta literalmente igual a la respuesta base dada por el docente para que la misma sea considerada como correcta. El prototipo realiza un recorrido tal por el grafo que respeta la libertad expresiva del alumno.

La operatoria del prototipo del sistema se describe a través de los siguientes pasos:

#### 4.1. Ingreso de preguntas

Como se mencionó en la sección anterior, inicialmente el docente confecciona un examen que está formado por un conjunto de preguntas asociadas, él como usuario docente, las ingresará en el sistema directamente mediante una interfaz gráfica de usuario (Graphic User Interface, GUI)[17, 18], las preguntas se almacenarán en la base de datos para luego asociarlas las respuestas correspondientes.

#### 4.2. Ingreso de respuestas

El ingreso del texto de las respuestas se realiza directamente mediante una GUI, el texto puede ser una respuesta base o candidata, el mismo se almacena de forma temporal para realizar todos los pasos siguientes.

Hay que mencionar que el tamaño del texto permitido tiene un límite definido empíricamente a partir de respuestas obtenidas en diferentes instancias de evaluación tales como exámenes parciales, exámenes finales, cuestionarios, etc. Este límite permite acotar el procesamiento de la respuesta con el consiguiente ahorro de tiempo pero a la vez permite la suficiente extensión como para que la respuesta sea lo suficientemente expresiva.

En el caso de las respuestas suministradas por los alumnos, el límite se establece en un 30 % más de caracteres que el establecido para las respuestas de los docentes. En todos los casos se presenta al usuario el tamaño máximo permitido y la cantidad de caracteres restantes que aún puede utilizar en la redacción de su respuesta.

#### 4.3. Corrección ortográfica y gramatical

Realiza la revisión ortográfica y sintáctica del texto. Para ello se realiza una revisión por palabras, contra un diccionario de idioma español y se verifican las reglas gramaticales básicas, antes de suministrar el texto al motor de búsqueda.

#### 4.4. Simplificación del texto

El texto se divide en unidades atómicas o tokens que se envían al motor de búsqueda para que el mismo determine si son conceptos o relaciones pre-existentes en el grafo. En todos los casos se trabaja con las raíces de los conceptos y los verbos en infinitivo.

El motor de búsqueda devuelve la misma lista de tokens con marcadores que indican si cada uno de los términos existe, no existe o es una equivalencia de un concepto existente (véase Cuadro 1). Si el mismo no existe será posible descartarlos de la consulta, reformularlos o solicitar que se agreguen como conceptos nuevos.

⇒	A	B	C	D	E	F	Vector inicial
⇐	A	∅	C	∅	∅	F	B,D,E no existen
⇒	A	B[+]	C	H		F	Agregar B, Reemplazar D por H, Descartar E
⇐	A	B	C	H		F	B Agregado, H existe
⇒	A	B	C	H		F	Vector final de búsqueda

Cuadro 1: Tabla de ajuste del vector de búsqueda

#### 4.5. Búsqueda de rutas

El objetivo de este paso es obtener de la base de datos orientada a grafos, el conjunto de todas las rutas que contienen los conceptos de la respuesta candidata.

#### 4.6. Valoración y ponderación de conceptos y relaciones

La búsqueda de conceptos y relaciones se utiliza para evaluar el grado de validez de la respuesta candidata.

Para ello todas las posibles rutas encontradas son comparadas con la respuesta base del docente y se computa el grado de validez de la respuesta candidata en base a esa comparación. El sistema asigna un valor a cada ruta encontrada, aplicando un algoritmo basado en la cantidad de conceptos y relaciones exactas que la componen y a la ponderación obtenida de cada concepto y relación de la respuesta candidata a partir de su comparación con la respuesta base.

Comparando las rutas encontradas con la respuesta candidata es posible detectar diferentes situaciones o casos que son detallados en [3] y que son contemplados en el prototipo del sistema.

#### 4.7. Evaluación de las respuestas

Al finalizar el cálculo de las ponderaciones, donde se compara la respuesta candidata del alumno con la respuesta base del docente, se obtiene un valor para la respuesta candidata, en el que podrá variar entre 0 y el valor de la respuesta base:  $V \leq V_{base}$ .

Un caso posible será que  $V = V_{base}$ , el cual implicará una respuesta correcta "perfecta" con un grado de coincidencia conceptual del 100 %.

Los valores intermedios indican el grado de aproximación de la respuesta candidata a la respuesta base y pueden ser útiles como indicadores para calcular el porcentaje de acierto en que la respuesta analizada está incluida.

En tal sentido es posible fijar dos umbrales ( $V_{min}$  y  $V_{max}$ ) a partir de los cuales se considera a las respuestas dentro de los siguientes parámetros:

$$V \geq V_{max} \rightarrow \text{correcta}$$

$$V \leq V_{min} \rightarrow \text{incorrecta}$$

$$V_{min} < V < V_{max} \rightarrow \text{evaluar}$$

Si el valor obtenido no se encuentra dentro de los extremos establecidos, la respuesta debe ser evaluada por un docente, a fin de asignar a la misma, una calificación correcta en base a una interpretación o un contexto que exceden aquellos representados por la base de conocimientos disponible, y por ende, al mecanismo de evaluación automatizado expuesto.

Por este motivo, el prototipo del sistema contempla un rango entre el  $V_{min}$  y  $V_{max}$ , donde requerirá la intervención del docente para que revise la respuesta del alumno. Esto es para los casos en que la respuesta no esté perfectamente calificada, pudiendo ocurrir dos situaciones:

- Que la respuesta del alumno sea correcta para el docente, pero los conceptos usados por el alumno no están en la base de conocimiento y por lo tanto se deberá, a discreción del docente, insertarlo como una nueva respuesta en la base de conocimiento y calificar nuevamente al alumno.
- Que la respuesta del alumno sea incorrecta para el docente, pero como utilizó algunos conceptos acertados y la combinación de los mismos le arrojó un rango de calificación aceptable para aprobar, será el docente el que decida cuál es el valor de calificación final de dicha respuesta.

## 5. TECNOLOGÍA UTILIZADA

La tecnología que se utilizó para crear el prototipo incluye herramientas de código abierto, ya que el uso principal se realizará en un ambiente universitario y por eso es de fundamental importancia no depender de ningún tipo de licenciamiento propietario. Asimismo era importante prever la posibilidad de implementación en otras unidades académicas por lo que la infraestructura de soporte no debía ser difícil de replicar.

Se hizo hincapié en herramientas y librerías que estuvieran en un estado apto para producción, o al menos en un estado beta avanzado, debido a que el eje de investigación no radicaba en el estudio de las herramientas sino en la utilización de las mismas para la implementación de una plataforma funcional.

Asimismo se prestó especial atención a las opiniones de la comunidad de usuarios, a la velocidad de respuesta de los desarrolladores frente a problemas y a consultas técnicas, a la frecuencia con que las herramientas recibían soporte y actualizaciones y al tiempo transcurrido desde la última actualización, como parámetros de calidad antes de tomar la decisión definitiva de qué herramientas y librerías utilizar.

### 5.1. Lenguaje de programación

Se realizó la implementación en lenguaje Java ya que permite implementar la arquitectura en multicapa y por la fácil intercambiabilidad de componentes que ofrece. La característica multiplataforma de Java, es muy importante en un proyecto de investigación relacionado con la educación universitaria, en donde es muy factible que distintas unidades académicas, posean distintas infraestructuras de hardware y software para implementar una solución de este tipo.

A esto debe agregarse la perfecta integración que tiene con las librerías de corrección ortográfica, la base de datos de grafos, y la librería de visualización, todas desarrolladas en este mismo lenguaje.

### 5.2. Librerías de corrección ortográficas

Para lograr un correcto análisis del texto, es importante que el mismo no posea errores del tipo ortográficos ni gramaticales. Al ser posible que un usuario del sistema ingrese un texto que no cumpla adecuadamente ambas características, surge la necesidad de brindar ayuda de contexto que permita identificar el tipo de falencias mencionadas en la redacción del mismo.

Para este fin se utilizó la herramienta LanguageTool que es un software que se encuentra bajo la licencia GNU Lesser General Public License(2.1), el cual brinda verificación gramatical y ortográfica. Está diseñado para que se pueda modificar su diccionario fácilmente, al igual que su conjunto de reglas gramaticales, logrando una gran flexibilidad en el momento de su implementación. El proyecto está preparado para ser utilizado en diversos tipos de aplicaciones lo que garantiza la portabilidad. Esta herramienta se adapta de mejor manera para llevar a cabo la ayuda de contexto (análisis ortográfico y sintáctico) que el análisis de texto requiere en esta etapa y permite una fácil integración en el lenguaje Java.

### 5.3. Base de datos de grafos

Para el manejo de la base de datos de grafos se utilizó el producto OrientDB Community Edition, es una aplicación de código abierto, con licencia Apache 2 y gratuita para todo tipo de uso. Esta base de datos de grafos implementa de forma nativa dos características que son centrales en el planteo del método de corrección.

- Nodos y Arcos etiquetados
- Tipos de datos complejos como atributos de los Nodos

Esas dos características hicieron posible que la implementación de los modelos teóricos planteados fuera directa, con el consiguiente ahorro en tiempos de desarrollo y simplicidad a la hora de utilizar dichos conceptos como parte del método de corrección.

### 5.4. Librería gráfica

La librería utilizada para realizar la visualización de la base de grafos fue GraphStream, que es una librería Open Source implementada en Java que provee toda la funcionalidad de visualización y trazado de rutas con una gran flexibilidad y facilidad de uso.

El prototipo hace un uso intensivo de las propiedades de ruteo disponibles en la librería GraphView de forma tal que la visualización es clara y con la menor cantidad de cruces de líneas entre los nodos. Un detalle interesante es que el algoritmo de distribución (layout) de la librería GraphStream es un módulo reemplazable (plugin) por lo que se puede definir un algoritmo distinto, de acuerdo a las necesidades del proyecto, y utilizarlo en lugar del layout por defecto.

## 6. CASO DE APLICACIÓN DE LA PROPUESTA

### 6.1. Preparación

Para realizar la prueba de concepto del prototipo, se diseñaron exámenes de prueba con el objetivo de comparar los resultados de las evaluaciones de los docentes de la cátedra de Paradigmas de Programación con los resultados de las evaluaciones del prototipo.

Los exámenes de prueba fueron diseñados por los docentes de la cátedra de Paradigmas de Programación y están compuestos por 10 (diez) preguntas sobre las seis unidades temáticas que contiene la materia y cuyas respuestas serán redactadas por los estudiantes, con escritura de texto libre.

El diseño de las preguntas se realizó de tal manera de que la respuesta candidata aportada por el alumno pudiera ser concreta y acotada. Además, para facilitar la respuesta del mismo, se asignó a cada pregunta un texto de inicio, por ejemplo a la pregunta: "Qué es un objeto?" se le asignó el texto de inicio: "Un objeto es: . . .".

El equipo de investigación se encargó de realizar el alta en el prototipo de las preguntas y de las respuestas base asociadas, de cargar las ponderaciones a los conceptos y a las relaciones enviadas por los docentes de cátedra, como así también los pesos correspondientes a sus equivalentes por cada respuesta base.

También se definió el rango de valores (ver Sección 4.7, Evaluación de las respuestas) que se utiliza para determinar si una respuesta candidata es correcta, incorrecta o si dicha evaluación necesita la intervención de un docente. Este rango, actualmente, se considera entre el 40 % y el 60 %. Si bien en estos casos el prototipo arroja una calificación, le sugiere al docente la intervención para que el mismo confirme la calificación. De esta manera, se considera que:

- Si el prototipo puntuó la respuesta del alumno con una valoración menor al 40 % del puntaje de la respuesta base se considera incorrecta
- Por otra parte si el prototipo puntuó la respuesta candidata con un valor superior al 60 % del valor de la respuesta base, se considera correcta
- Finalmente si el prototipo puntuó la respuesta del alumno con una valoración entre el 40 % y el 60 % del puntaje de la respuesta base se considera que requiere la intervención del docente

De acuerdo a la cantidad de caracteres de las respuestas base dada por el docente, el prototipo del sistema asigna a cada pregunta una cantidad máxima de caracteres que se espera obtener de respuesta del alumno.

Se ha considerado un 30 % más de caracteres que la máxima cantidad que contiene la respuesta base del docente. Este porcentaje fue determinado considerando las respuestas que los alumnos han escrito en exámenes tradicionales, en papel, y al contenido del material de estudio desarrollado por la cátedra. En tal sentido cabe aclarar que la cátedra de Paradigmas de Programación dispone de un único material de estudio para todos los cursos.

Actualmente los exámenes tanto parciales y finales de la materia son realizados de manera unificada, es decir que se evalúa a todos los alumnos el mismo día y con el mismo instrumento de evaluación.

Esta evaluación tiene en cuenta la escala de notas que se utiliza actualmente en el Departamento de Ingeniería en Sistemas de Información, Universidad Tecnológica Nacional, Facultad Regional Córdoba y se que muestra a continuación:

Nota	Porcentaje	Calificación
1		Insuficiente
2		Insuficiente
3		Insuficiente
4		Insuficiente
5		Insuficiente
6	60 % a 68 %	Aprobado
7	69 % a 77 %	Bueno
8	78 % a 86 %	Muy bueno
9	87 % a 95 %	Distinto
10	96 % a 100 %	Sobresaliente

Cuadro 2: Escala de notas

De acuerdo a esta escala, se califican las respuestas candidatas que hayan obtenido una calificación menor al 60 % como incorrectas, y para las que hayan tenido valores iguales o superiores al 60 % se califican como correctas.

En última instancia se totalizan todas las respuestas obtenidas para obtener la evaluación del examen en su totalidad.

## 6.2. Ejecución

Para la ejecución de la prueba, se tomó el examen a los alumnos pertenecientes a una misma comisión o curso. De la evaluación obtenida, se consideró una muestra representativa de exámenes con respuestas variadas por parte de los alumnos, totalizando cincuenta respuestas a calificar.

Se seleccionaron cinco docentes de la cátedra de Paradigmas de Programación para que realizaran la corrección de los exámenes en forma manual y propusieran un puntaje para cada pregunta, un valor entre 0 y 10, y una nota final al examen de acuerdo a la escala de notas mencionada en el punto anterior.

## 6.3. Resultados

Comparado con la longitud máxima especificada para las respuestas escritas por los alumnos, se evidencia que las respuestas candidatas obtenidas se encuentran todas dentro del 90 % de los caracteres asignados a su correspondiente respuesta base, por lo que se concluye que el tamaño inicialmente previsto es correcto y permitiría un buen grado de expresividad en las respuestas.

A continuación, se presentan los resultados obtenidos analizando la totalidad de las evaluaciones realizadas y seleccionando para presentar los ejemplos tres de las evaluaciones.

Cada uno de ellas presenta la pregunta que se evaluó y todas las respuestas candidatas dadas por los alumnos, denominadas como  $T_1 \dots T_n$ . Para ellas se consideraron las calificaciones asignadas manualmente por los docentes, a través de los valores máximo, mínimo y promedio (nota promedio entre todas las calificaciones obtenidas).

Estos valores se compararon con la calificación obtenida automáticamente por el prototipo del sistema, referenciado en las tablas como Sibila<sup>1</sup>.

1. Sibila: Nombre de la primera pitonisa del Oráculo de Delfos a la que los viajeros acudían por "respuestas" y guía.

Se presenta además, para cada pregunta, la tabla con los valores máximo, mínimo y promedio de los docentes, así como la calificación obtenida de la evaluación automatizada que provee Sibila. En base a los datos tabulados se muestra un gráfico con los valores obtenidos para una mejor visualización.

Se muestran a continuación los datos y gráficos resultantes del análisis de las preguntas mencionadas.

### 6.3.1. ¿Qué es un paradigma de programación?

- $T_1$  Un modelo básico de diseño e implementación de programas
- $T_2$  Un conjunto de opciones
- $T_3$  Aplicaciones en assembler
- $T_4$  Un entorno de desarrollo
- $T_5$  Un enfoque particular o filosofía para la construcción del software

	MAX	MIN	AVG	SIBILA
$T_1$	10	10	10	10
$T_2$	2	0	0.6	0
$T_3$	1	0	0.2	0
$T_4$	4	0	0.8	0
$T_5$	10	8	9.6	8

Cuadro 3: Calificaciones

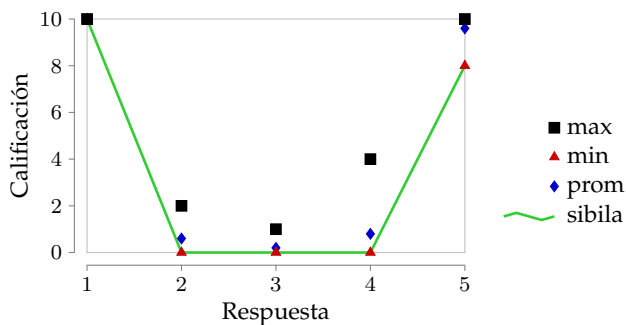


Figura 3: P1 - ¿Qué es un paradigma de programación?

### 6.3.2. ¿Cómo se define un programa en POO?

- $T_1$  Un conjunto de objetos que colaboran entre sí enviándose mensajes
- $T_2$  Un conjunto de objetos donde los objetos responden a pedidos interactuando con los otros objetos que conoce
- $T_3$  Un conjunto de objetos que colaboran entre sí enviándose mensajes
- $T_4$  Un conjunto de objetos que colaboran entre sí enviándose mensajes
- $T_5$  Un conjunto de objetos que envían mensajes

	MAX	MIN	AVG	SIBILA
$T_1$	10	10	10	10
$T_2$	10	8	9.6	10
$T_3$	10	10	10	10
$T_4$	10	10	10	10
$T_5$	10	7	9	10

Cuadro 4: Calificaciones

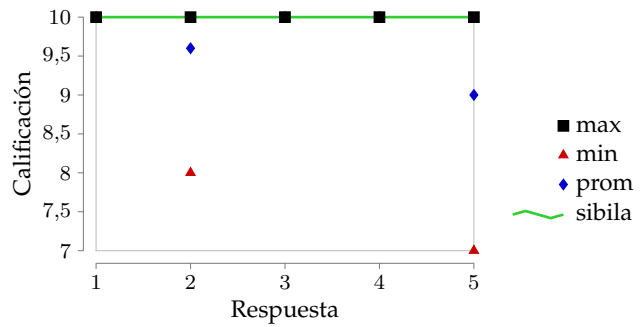


Figura 4: P2 - ¿Cómo se define un programa en POO?

### 6.3.3. Cuáles son los campos de aplicación del paradigma lógico?

- $T_1$  Inteligencia Artificial, Sistemas basados en el conocimiento, Procesamiento del lenguaje natural
- $T_2$  Inteligencia Artificial, Sistemas basados en el conocimiento, Sistemas donde un sistema de información imita las recomendaciones de un experto sobre algún dominio de conocimiento
- $T_3$  Sistemas expertos, Sistemas basados en el conocimiento, Procesamiento del lenguaje natural
- $T_4$  Inteligencia Artificial, Demostración automática de teoremas, Procesamiento del lenguaje abstracto
- $T_5$  Sistemas expertos, Sistemas basados en el conocimiento, Procesamiento del lenguaje natural, Reconocimiento de lenguaje natural

	MAX	MIN	AVG	SIBILA
$T_1$	10	5	8.8	10
$T_2$	10	5	8.6	10
$T_3$	10	6	9	9.7
$T_4$	10	4	6.8	7.5
$T_5$	10	5	8.2	10

Cuadro 5: Calificaciones

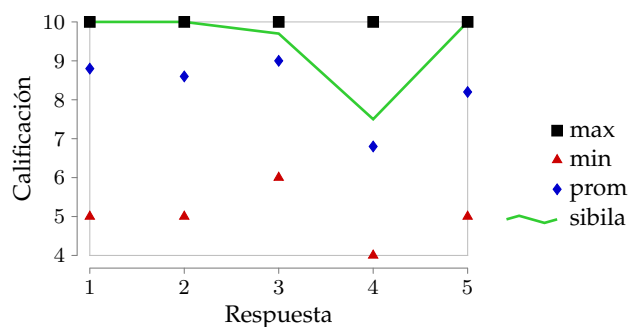


Figura 5: P5 - ¿Cuáles son los campos de aplicación del paradigma lógico?

## 6.4. Análisis estadístico de los resultados

Se han realizado algunos cálculos estadísticos descriptivos para determinar en qué grado el prototipo del sistema Sibila es consistente con la corrección manual de docentes experimentados.

Los resultados se resumen en la siguiente tabla:

Concepto	Docentes	Sibila
Nota promedio	Entre 6.93 y 7.73	7.68 (1)
Desviación estándar	Entre 3.488757 y 4.067291	4.049727 (2)
Curtosis	Entre -1.0217 y 0.863368	0.608147
Asimetría	Entre -1.56357 y -0.75183	-1.53663 (3)

Cuadro 6: Caracterización de los resultados

Análisis:

1. Dentro del rango, la media es coincidente con el puntaje dado por los docentes en su corrección
2. Dentro del rango, desviación coincidente con la calificación dada por los docentes
3. Dentro del rango, asimetría izquierda, es decir que los docentes califican generalmente por encima del valor medio (5, en esta escala de notas). Esto es coherente con la media aritmética, valuada entre 6.93 y 7.73

El siguiente gráfico muestra la distribución estadística de las calificaciones colocadas por los docentes, en todo el rango de preguntas utilizadas como muestra, y la distribución de calificaciones colocadas por el sistema Sibila.

Se observa que el sistema arroja valores similares a los que asignan los docentes al calificar una pregunta.

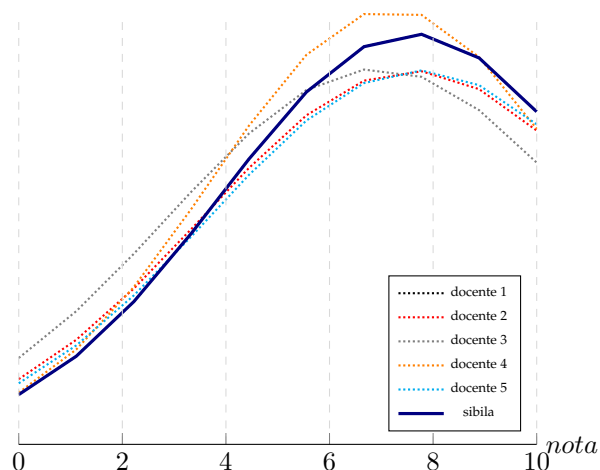


Figura 6: Correlación entre Sibila y los docentes de muestra

### 6.5. Apreciación de los docentes

Los docentes que participaron en las pruebas del prototipo, si bien en esta primera etapa no lo usaron directamente, manifestaron la conformidad con los resultados obtenidos en el proceso de calificación e indicaron que es una herramienta innovadora para utilizar en la cátedra y están dispuestos a hacer las pruebas en cursos modelos para verificar completamente la confiabilidad y viabilidad del prototipo propuesto.

En general los docentes manifestaron conformidad sobre implementar una herramienta automática que permita evaluar, no solo los conocimientos requeridos en la cátedra, sino también la capacidad expresiva de las respuestas de los alumnos.

Todos concuerdan que la herramienta de evaluación automática propuesta, podría permitir lo siguiente:

- Disminuir los tiempos de corrección de los trabajos por los docentes (de modo que los errores puedan ser detectados y, si es necesario, adaptar ciertos contenidos).
- Aumentar la base de conocimiento de respuestas, a partir de nuevas definiciones que el alumno puede introducir y que el docente puede validar como ciertas.
- Fomentar la expresividad de los alumnos en las respuestas que redacte.
- Introducir un ambiente de trabajo más amigable para los alumnos.
- Favorecer el feedback para los alumnos que les interesa recibir la retroalimentación ya que enlazan su producción con los resultados.
- Posibilidad de emitir distintos tipos de informes: individuales para cada alumno, por cohortes, informes psicométrico por ítem y prueba, etc.

## 7. CONCLUSIONES Y TRABAJOS FUTUROS

Se ha presentado el prototipo de un sistema de corrección automatizado, que permite calificar exámenes formados por preguntas a ser respondidas por alumnos de nivel universitario, a través de redacción en texto libre, en un dominio acotado.

En base a los resultados de la implementación en un entorno de prueba, se ha demostrado que el prototipo ha podido evaluar un conjunto de respuestas en un nivel similar a como lo haría un docente.

El prototipo provee, además, la posibilidad de mejorar sensiblemente los tiempos de corrección y brindar un mecanismo más objetivo a la hora de evaluar los conocimientos de los alumnos, pero sin perder la posibilidad de intervención del docente para ajustar los criterios y servir como elemento de realimentación del sistema a fin de que el mismo mejore los resultados que entrega.

La percepción de los docentes involucrados en el uso del prototipo, ha sido muy positiva, los mismos han manifestado que es un avance significativo en la corrección automatizada, respecto de la herramienta de múltiple opción usada actualmente, ya que permite evaluar la expresividad del alumno, sin dejar de demostrar la comprensión de los conceptos básicos. Según los mismos docentes, estarían dispuestos a utilizar esta herramienta como parte del proceso de corrección.

Actualmente, se está trabajando en una segunda etapa sobre una plataforma web, que cambiará la vista de la interface, para permitir, tanto a los docentes como a los alumnos, interactuar con el sistema de una manera más simple y desde diferentes lugares de trabajo.

Paralelamente, se continúa con el análisis de textos con el objeto de completar la base de grafos representativa del dominio de conocimiento de la materia, y el reconocimiento de respuestas que contengan mayor cantidad de palabras y de mayor complejidad.

Es importante mencionar que al utilizar el sistema e ir completando la base de conocimientos se pueden obtener importantes indicadores con respecto al dictado de la materia, al propio instrumento de evaluación, así como a los métodos de enseñanza utilizados[19].



## REFERENCIAS

Estos indicadores pueden ser utilizados posteriormente para ajustar tanto los contenidos de la materia como la forma y/o contenidos dictados y evaluados.

Se espera que, con esta herramienta, los conocimientos y habilidades adquiridas sean realmente significativas y eleven el rendimiento en el cursado y aprobación de la asignatura, y sea posible extender su aplicación a otras cátedras de la carrera Ingeniería en Sistemas de Información que se dictan en la Facultad Regional Córdoba de la Universidad Tecnológica Nacional.

### 7.1. Trabajos futuros

Con respecto a las pruebas, está previsto extender la muestra al resto de las comisiones en que se dicta la materia, comenzando por un segundo curso y finalizando con toda la cátedra, para validar la premisa básica del trabajo con respecto a la calidad de la evaluación realizada automáticamente por el sistema así como la ponderación realizada por el mismo sobre los conceptos y para la obtención de la nota final del examen, teniendo en cuenta los resultados obtenidos en esta primera prueba de concepto. En este caso, se realizaría con la nueva interface web implementada.

También se planifica llevar a cabo una encuesta tanto para docentes como alumnos para conocer los beneficios y desventajas que encontraron al momento de la configuración y uso del prototipo. Dicha información permitirá mejorar no solo las cuestiones de usabilidad y la carga del prototipo sino también detectar falencias en la libertad expresiva que se pretende lograr.

"[...] numerosos autores (Hammond y Collins, 1991; Reeves y Okey, 1996) proponen el diseño de entornos evaluativos constructivistas con un fuerte apoyo de las Nuevas Tecnologías de la Información (NTI): formularios en HTML que permiten cierto feedback al profesor, software con pruebas cerradas que permiten autocorrección por parte de los alumnos, a la vez que facilitan la comunicación a distancia, la autodirección, la autorregulación y una mayor dinámica del proceso evaluativo." [20].

Atentos a ello, proponemos el uso del prototipo por parte de los alumnos en instancias de autoevaluación. Que desde sus hogares, y respetando sus tiempos de estudio, puedan disponer del prototipo para autoevaluar el avance de su aprendizaje. Brindar a los docentes la posibilidad de llevar a cabo un seguimiento de sus alumnos incorporando en el prototipo datos estadísticos que permitan ajustar la formulación de las preguntas, y sobre todo incorporar respuestas dadas por los alumnos no contempladas por el momento.

Esto permitirá enriquecer la base de conocimiento. Así también, ".. permitirá entre otras cosas, poder evitar que los alumnos rechacen los exámenes; viéndolos como instrumentos exclusivamente comprobatorios, formales o como un pase de cuentas. Llegando a reconocer que ellos son una herramienta necesaria para su autorregulación, para ajustar lo aprendido, para retroalimentarse, para perfeccionar sus métodos de estudio e impulsar su aprendizaje hacia planos superiores" [20].

## REFERENCIAS

- [1] Marcelo Marciszack y col. *Material de estudio de la Cátedra Paradigmas de Programación*. Editado por Cátedra de Paradigmas de Programación Ingeniería en Sistemas de Información. Universidad Tecnológica Nacional, Facultad Regional Córdoba, 2016 (véase página 1).
- [2] María Alejandra Paz Menvielle y col. "Metodología para determinar la exactitud de una respuesta, escrita en forma textual, a un interrogante sobre un tema específico, aplicando herramientas informáticas". En: *II Congreso Nacional de Ingeniería Informática y Sistemas de Información*, CONAIIISI, Universidad Nacional de San Luis, Argentina (2014). 2014 (véase página 1).
- [3] María Alejandra Paz Menvielle y col. "Detección de conceptos y relaciones para evaluación de respuestas". En: *III Congreso Nacional de Ingeniería Informática y Sistemas de Información*, CONAIIISI, UTN, Facultad Regional Buenos Aires, Argentina (2015). 2015 (véanse páginas 1, 4).
- [4] María Alejandra Paz Menvielle y col. "Arquitectura y operatoria de un sistema de corrección de exámenes automatizado, utilizando grafos dirigidos". En: *IV Congreso Nacional de Ingeniería Informática y Sistemas de Información*, CONAIIISI, Universidad Católica de Salta, Facultad de Ingeniería, Argentina (2016). 2016 (véase página 1).
- [5] José Manuel García Ramos. *Bases pedagógicas de la evaluación: guía práctica para educadores*. 1989 (véase página 1).
- [6] Joan Mateo y Joan Mateo Andrés. *La evaluación educativa, su práctica y otras metáforas*. Volumen 33. HORSORI EDITORIAL, SL, 2000 (véase página 1).
- [7] M<sup>a</sup> Rodríguez Conde. "Aplicación de las TIC a la evaluación de alumnos universitarios". En: *Teoría de la Educación. Educación y Cultura en la Sociedad de la Información* 6.2 (2005) (véanse páginas 1, 2).
- [8] J Mateo. "La toma de decisiones en el contexto de la evaluación educativa". En: *Revista de investigación educativa* 8.16 (1990), páginas 95-112 (véase página 1).
- [9] JM García Ramos y R Pérez Juste. "Diagnóstico, evaluación y toma de decisiones". En: *Rialp. Madrid* (1989) (véase página 1).
- [10] Beatriz Pérez Sánchez y Flora Salas Madriz. "Hallazgos en investigación sobre el profesorado universitario y la integración de las TIC en la enseñanza". En: *Revista Electrónica "Actualidades Investigativas en Educación"* 9.1 (2009) (véase página 2).
- [11] Begoña Gros e Ingrid Noguera. "Mirando el futuro: evolución de las tendencias tecnopedagógicas en educación superior". En: *Revista Científica de Tecnología Educativa* 2.2 (2013), páginas 130-140 (véase página 2).
- [12] John F Sowa. "Conceptual graphs for a data base interface". En: *IBM Journal of Research and Development* 20.4 (1976), páginas 336-357 (véase página 2).
- [13] W Hartshorne. *Burks, editors. Collected Papers of Charles Sanders Peirce, Cambridge, Massachusetts, 1931-1935* (véase página 2).
- [14] Frank Van Harmelen, Vladimir Lifschitz y Bruce Porter. *Handbook of knowledge representation*. Volumen 1. Elsevier, 2008 (véase página 2).

- [15] Ivan Olmos, Jesus A Gonzalez y Mauricio Osorio. "Inexact Graph Matching: A Case of Study." En: *FLAIRS Conference*. 2006, páginas 586-591 (véase página 2).
- [16] J Glenn Brookshear. *Theory of computation: formal languages, automata, and complexity*. Benjamin-Cummings Publishing Co., Inc., 1989 (véase página 2).
- [17] HK Tabbers y col. "Interface design for digital courses". En: *Integrated E-learning: Implications for pedagogy, technology and organization* (2004), páginas 100-111 (véase página 4).
- [18] Vincent Cho, TC Edwin Cheng y WM Jennifer Lai. "The role of perceived user-interface design in continued usage intention of self-paced e-learning tools". En: *Computers & Education* 53.2 (2009), páginas 216-227 (véase página 4).
- [19] Gloria Patricia, Sandra Cristina y col. "Propuesta para la medición del impacto de las TIC en la enseñanza universitaria". En: (2011) (véase página 8).
- [20] M Hernández Nodarse. "Perfeccionando los exámenes escritos: reflexiones y sugerencias metodológicas". En: *Revista Iberoamericana de educación* 41.4 (2007), páginas 1-25 (véase página 9).