

Model and evaluation tool using graphs as knowledge base for the automated correction of exams in text format

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*



Abstract—The present work describes a method to analyze the answers of questions of an examination of the class Paradigms of Programming, given by students and written in the form of text written in natural language, to contrast their degree of coincidence with the answers to the same questions that were provided by the teachers of the class.

From the proposal arise elements called concepts and / or relations to which values will be assigned in order to be able to weight the response provided by the student and compare it with the weighting of the base answers elaborated by the teachers.

The techniques presented consider all possible situations that may arise in the student's response. As part of the field work, a sample of questions was answered by the students, corrected and qualified by teachers of the class Programming Paradigms, in order to be compared with the results of the prototype of the constructed system, whose results are exposed in this work.

Index Terms—text analysis; graph; pattern detection; recognition; path detection

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.

2. INTRODUCCIÓN

El procesamiento de patrones 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.

En éste ámbito, se realiza el estudio sobre la representación del conocimiento utilizando grafos conceptuales.

Un grafo conceptual[2] es un sistema de notación simbólica y de representación del conocimiento. Presentado por John F. Sowa, se basa en los gráficos existenciales[3] de Charles Sanders Pierce, en las estructuras de redes semánticas y en datos de la lingüística, la filosofía y la psicología[4].

El enfoque elegido para realizar la búsqueda de patrones dentro de la información almacenada en forma de grafo, es el de búsquedas inexactas, ya que permite determinar la validez total o parcial, de las respuestas obtenidas. El algoritmo utilizado requiere especificar la estructura que se desea encontrar y 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[5].

La elección del enfoque de búsqueda de patrones, se realizó considerando la naturaleza inherentemente variable de las respuestas de los alumnos, registradas en exámenes al mismo conjunto de preguntas.

Actualmente el equipo de investigación ha desarrollado el prototipo de un sistema que trabaja con la búsqueda de patrones en una base de datos de grafos que 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[6].

Trabajos presentados sobre el uso de las tecnologías de información y comunicaciones (TIC) en el ámbito universitario ponen de manifiesto que si bien la implementación de dichas tecnologías provee resultados indiscutiblemente beneficiosos[7] 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”[8], 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 presentará un método 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.

El objetivo del presente trabajo es describir el mecanismo de corrección diseñado, utilizando grafos como base de conocimiento para la corrección automatizada de exámenes escritos por estudiantes en formato de texto y mostrar los resultados preliminares 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.

3. FUNCIONAMIENTO DEL PROTOTIPO

El prototipo del sistema de corrección automatizada de exámenes posee la arquitectura y la operatoria necesarias para registrar las preguntas de un examen y analizar las respuestas escritas por alumnos, en forma de texto redactado en lenguaje natural.

En la operatoria descrita se definen algunas entidades que son de uso general en todos los módulos involucrados y que sirven para modelar el funcionamiento de cada uno de ellos. Estas entidades son:

- **Término:** es el equivalente a una palabra en el texto de la respuesta que contiene información de gestión asociada al modelo de dominio. Todos los Términos están compuestos por una y sólo una palabra.
- **Concepto:** es un nodo que representa una palabra almacenada en la base de conocimiento, que tiene una equivalencia directa con uno o más conceptos de la materia Paradigmas de Programación.
- **Concepto Compuesto:** está formado por más de una palabra o término. A los fines de la evaluación el Concepto Compuesto se trata como una unidad indivisible.
- **Relación:** es un arco etiquetado, formado por una palabra, que une dos conceptos.

- **Relación Compuesta:** es un arco etiquetado, formado por más de una palabra, que une dos conceptos. La relación compuesta se utiliza de forma equivalente a la relación simple.
- **Estructura:** es la mínima unidad que se debe procesar para una evaluación. Está compuesta de dos Conceptos unidos por medio de una Relación, es el conjunto Concepto-Relación-Concepto (ver Figura 1). Hay que mencionar que una Relación, Simple o Compuesta puede unir dos Conceptos, Simples o Compuestos, en cualquier orden.

La estructura descrita se fundamenta en la teoría de grafos[9, 10], la cual establece que cada nodo contiene un conjunto de etiquetas diseñado para representar correctamente el concepto asociado y facilitar la posterior búsqueda.

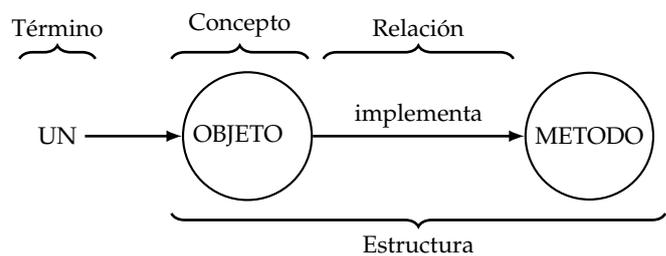


Figura 1: Estructura

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

3.1. Ingreso del texto

Inicialmente se realiza el ingreso del texto al sistema, directamente mediante una interfaz gráfica de usuario (Graphic User Interface, GUI)[11, 12] y dicho texto se almacena de forma temporal para realizar todas las actividades mencionadas a continuación.

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.

3.2. Corrección ortográfica y gramatical

En este paso se realiza la revisión ortográfica y sintáctica del texto. Para ello se efectúa 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.

3.3. 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	B	C	D	E	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

3.4. Búsqueda de rutas

Con los valores contenidos en el vector de búsqueda, se obtiene de la base de datos orientada a grafos, el conjunto de todas las rutas que contienen los conceptos coincidentes con los de la respuesta candidata.

3.5. 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 el siguiente apartado y que son contemplados en el prototipo del sistema.

4. EL PROCESO DE CORRECCIÓN

4.1. Técnicas y mecanismos de ponderación

A continuación se muestran las técnicas que permiten asignar valores a los conceptos y relaciones a fin de poder ponderar la respuesta suministrada por el alumno y compararla con la ponderación de la respuesta base elaborada por el docente.

Estas técnicas consideran todos los casos, incluyendo los distintos grados de acierto que pueda tener la respuesta del alumno, exponiendo los mecanismos con los que se deben analizar los conceptos y las relaciones para obtener la ponderación de la respuesta provista.

Caso 1: Si la respuesta candidata tiene los mismos conceptos y relaciones, en las mismas ubicaciones y con los mismos nombres que en la respuesta base, se la denomina "respuesta perfecta" y tendrá un valor que se calcula como:

$$V_r = C + R \quad (1)$$

Siendo:

V_r	Valor de la respuesta
C	Cantidad de conceptos
R	Cantidad de relaciones

En una respuesta perfecta se pondera cada concepto y relación con el valor 1. Cualquier otro caso implica que hay conceptos o relaciones inexactas (con valores ponderados < 1) y se requiere un análisis diferente de acuerdo a cada situación.

Seguidamente, se describen los distintos casos que se pueden presentar cuando la respuesta no es perfecta.

Caso 2: Los conceptos son coincidentes pero las ubicaciones dentro de la respuesta candidata difieren de la respuesta base (ver Figura 2).

En este caso se calcula, para cada concepto de la respuesta candidata, el desplazamiento con respecto a la respuesta base[13]. Dicho desplazamiento se define mediante dos componentes.

El primero es un valor numérico que expresa la diferencia entre las posiciones del concepto en la respuesta base y su ubicación en la respuesta candidata.

El segundo componente es un signo, que indica la dirección del desplazamiento. Por convención, se utiliza el signo positivo (+) para indicar conceptos que están desplazados hacia la derecha con respecto a la respuesta inicial y un signo negativo (-) para indicar aquellos conceptos que aparecen a la izquierda en la respuesta candidata que en la base.

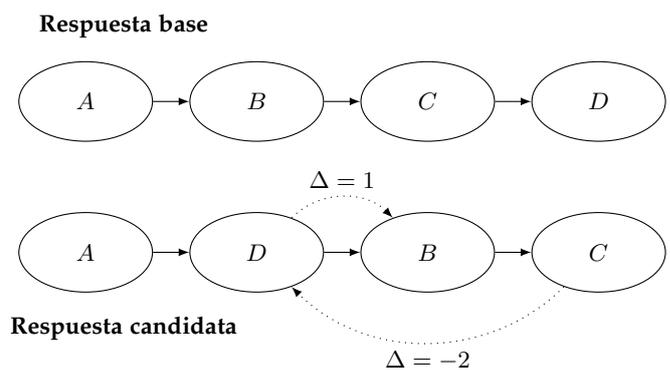


Figura 2: Desplazamiento entre conceptos

Se considera que mientras más alejado esté el concepto de su ubicación correcta, su valoración en la respuesta disminuye. Es por ello que el valor del concepto se pondera de la siguiente forma:

$$C_d = 1 - \left| \frac{P_c - P_b}{n} \right| \quad (2)$$

Siendo:

C_d	Valor del concepto desplazado
-------	-------------------------------

P_c Posición del concepto en la respuesta candidata
 P_b Posición del concepto en la respuesta base
 n Cantidad de conceptos en la respuesta base

Caso de aplicación (ver Figura 3, Cuadro 2):

Pregunta: ¿Qué es un objeto?

Respuesta base: Un objeto es una: “abstracción que tiene métodos y atributos”

Respuesta candidata: Un objeto es una: “abstracción que tiene atributos y métodos”

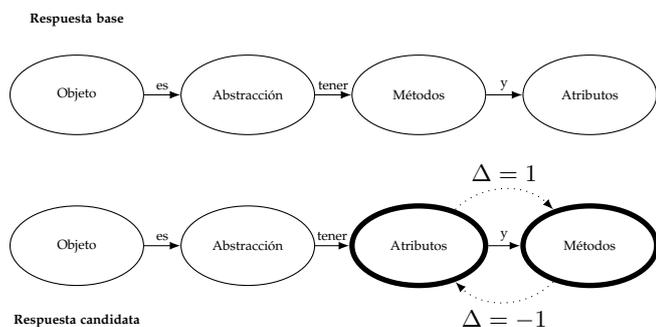


Figura 3: Conceptos desplazados. Caso de aplicación.

Los conceptos, en la respuesta candidata, se ponderan de la siguiente forma:

$$C_{metodos} = 1 - \left| \frac{4_c - 3_b}{4} \right| = 0,75$$

$$C_{atributos} = 1 - \left| \frac{3_c - 4_b}{4} \right| = 0,75$$

Concepto	Peso
Objeto	1
Abstracción	1
Atributos	0.75
Métodos	0.75

Cuadro 2: Ponderación de conceptos

Caso 3: Los conceptos son equivalentes a los conceptos de la respuesta base (ver Figura 4).

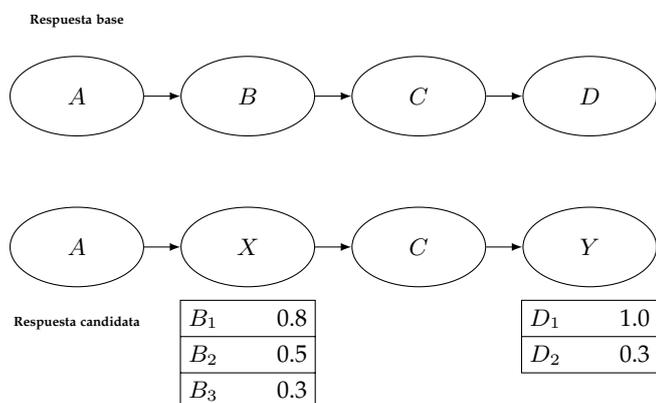


Figura 4: Peso de conceptos equivalentes

En este caso, cada concepto tendrá una tabla de equivalencias, definida por el docente, cuyo fin será ponderar cada una de las equivalencias a un valor numérico entre 0 y 1, donde 1 será una equivalencia completa, que indicará que ambos términos pueden usarse indistintamente. De esta forma el valor del concepto se calcula como:

$$C_e = C_{E_i} \quad (4)$$

Siendo:

C_e Valor del concepto equivalente
 C_{E_i} Valor de la equivalencia ponderada, según la respuesta base (ver Figura 4)

Caso de aplicación (ver Figura 5, Cuadros 3 y 4):

Pregunta: ¿Qué es un objeto?

Respuesta Base: Un objeto es una: “abstracción que utilizaremos como una representación computacional de una entidad de la realidad”

Respuesta Candidata: Un objeto es una: “abstracción que utilizaremos como una representación de una cosa”.

Concepto	Equivalencia	Peso
Entidad de la realidad	Ente	0.7
Entidad de la realidad	Cosa	0.3
Entidad de la realidad	Elemento	0.2
Representación computacional	Representación	0.4

Cuadro 3: Tabla de equivalencias

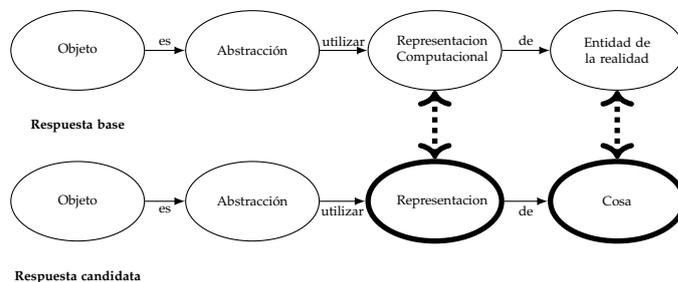


Figura 5: Conceptos equivalentes. Caso de aplicación.

Concepto	Peso
Objeto	1
Abstracción	1
Representación	0.4
Cosa	0.3

Cuadro 4: Ponderación de conceptos

Los casos 1 y 2 descritos se podrán combinar cuando existan conceptos equivalentes, y además se encuentren en ubicaciones diferentes a las definidas en la respuesta base, ponderando el concepto candidato según:

$$C_{de} = C_d \cdot C_e \quad (5)$$

Siendo:

C_{de} Valor del concepto desplazado equivalente

C_d Valor del concepto desplazado
 C_e Valor del concepto equivalente

Caso 4: La relación es exacta pero un nodo o ambos, nodo origen y nodo destino, no son exactos ni equivalentes. En este caso el peso de la relación se reduce, debido a que no hay forma de asegurar que la relación original es válida, al haber cambiado la exactitud de los nodos asociados. En este caso el valor de la relación se calcula como:

$$V_r = \frac{(1 + C_o + C_d)}{3} \quad (6)$$

Siendo:

V_r Valor de la relación
 C_o Peso del concepto de origen
 C_d Peso del concepto de destino

Como se puede observar a través de la fórmula, el caso de la relación exacta pura es un caso particular en el que el peso del concepto origen y del concepto de destino son ambos igual a 1.

Caso de aplicación (ver Figura 6 y Cuadro 5):

Pregunta: ¿Qué es un objeto?

Respuesta Base: Un objeto es una: “abstracción que utilizaremos como una representación computacional de una entidad de la realidad”

Respuesta Candidata: Un objeto es una: “abstracción que utilizaremos como una representación computacional de un programa”.

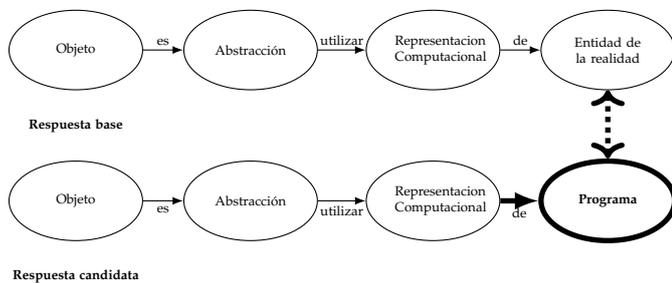


Figura 6: Ponderación de la relación de entre representación computacional y programa

En este caso, el concepto “programa” no existe, lo que explica la ponderación en 0, en la fórmula 7

$$V_r = \frac{(1 + 1 + 0)}{3} = 0,67 \quad (7)$$

Concepto	Peso
Es	1
Utilizar	1
De	0.67

Cuadro 5: Ponderación de relaciones

Caso 5: Los conceptos son exactos y la relación es un término equivalente. En este caso se debe analizar la lista de equivalencias suministrada por el sistema. Esta lista

contiene los términos equivalentes con su peso asociada para una relación dada en la respuesta base.

Si la relación en la respuesta candidata se encuentra dentro de la tabla de equivalencias, el valor R_e de dicha relación, se obtiene directamente de la tabla.

Caso 6: La relación es equivalente y los nodos de origen y/o destino también son equivalentes. En este caso se utiliza la tabla de equivalencias y el peso de la relación es alterado de acuerdo al peso equivalente del nodo origen y/o destino, y el de su propio peso equivalente (ver Figura 7).

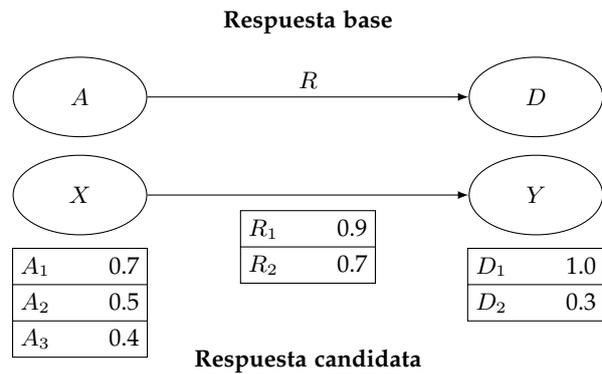


Figura 7: Relaciones y conceptos equivalentes

Aquí el valor de la relación se define como la media matemática entre los valores de los nodos de origen y de destino y el valor ponderado de la relación equivalente, como se indica en la fórmula 8.

$$V_r = \frac{R_e + C_o + C_d}{3} \quad (8)$$

Siendo:

V_r Valor de la relación
 R_e Valor de la relación equivalente
 C_o Valor del concepto de origen
 C_d Valor del concepto de destino

En el ejemplo de la figura 7, la relación

$$A_1 \rightarrow R_1 \rightarrow D_2$$

tendría un valor ponderado de

$$V_r = \frac{(A_1 + R_1 + D_1)}{3}$$

$$V_r = \frac{(0,7 + 0,9 + 0,3)}{3} = 0,633$$

Caso de aplicación (ver Figura 8 y Cuadros 6, 7 y 8):

Pregunta: ¿Qué es un objeto?

Respuesta Base: Un objeto es una: “abstracción que utilizaremos como una representación computacional de una entidad de la realidad”

Respuesta Candidata: Un objeto es una: “abstracción que utilizaremos como una representación para entidades”.

Concepto	Equivalencia	Peso
Entidad de la realidad	Ente	0.7
Entidad de la realidad	Cosa	0.3
Entidad de la realidad	Elemento	0.2
Representación computacional	Representación	0.4

Cuadro 6: Tabla de equivalencias. Conceptos.

Relación	Equivalencia	Peso
de	para	0.7

Cuadro 7: Tabla de equivalencias. Relaciones.

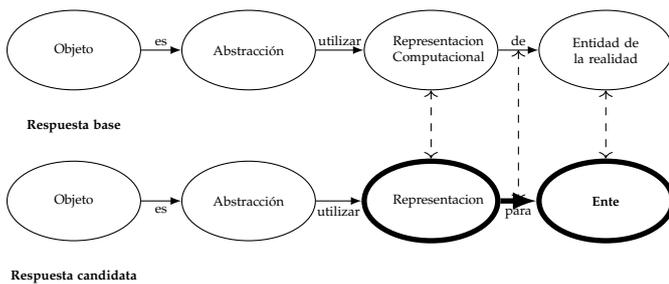


Figura 8: Relaciones y conceptos equivalentes. Caso de aplicación.

El valor real de la relación “para” se calcula como:

$$V = \frac{(R_e + C_o + C_d)}{3}$$

$$V = \frac{(0,7 + 0,4 + 0,7)}{3} = 0,6$$

Relación	Peso
Es	1
Utilizar	1
Para	0.6

Cuadro 8: Ponderación de relaciones

4.2. 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 en los casos mencionados en la sección 4.1, 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. CASO DE APLICACIÓN DE LA PROPUESTA

5.1. Preparación

Para realizar la prueba de concepto del prototipo, se propuso verificar los casos previamente descritos en un entorno simulado, para ello 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 sea 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 4.2, 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 %	Distinguido
10	96 % a 100 %	Sobresaliente

Cuadro 9: 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.

5.2. 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.

5.3. 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 tres de las evaluaciones realizadas, a modo de ejemplo y para analizar los resultados obtenidos.

Cada uno de ellas presenta la pregunta que se evaluó y todas las respuestas candidatas, 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 Sibila¹.

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.

5.3.1. ¿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 10: Calificaciones

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

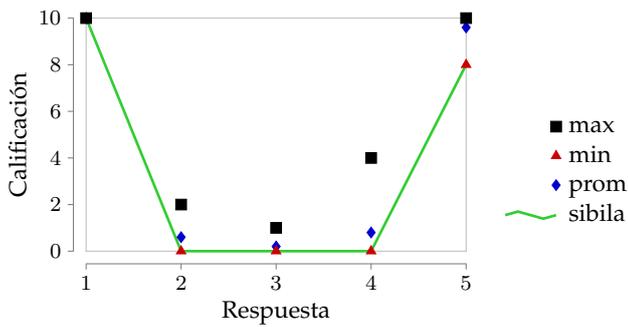


Figura 9: P1 - ¿Qué es un paradigma de programación?

5.3.2. ¿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 11: Calificaciones

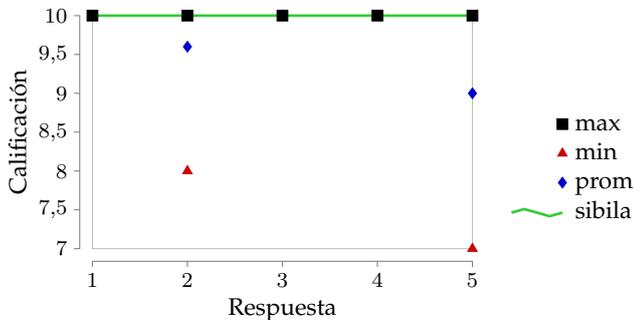


Figura 10: P2 - ¿Cómo se define un programa en POO?

5.3.3. 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 12: Calificaciones

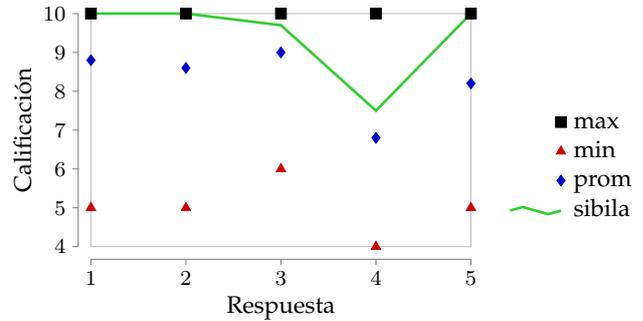


Figura 11: P5 - ¿Cuáles son los campos de aplicación del paradigma lógico?

5.4. 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 13: Caracterización de los resultados

Análisis:

- Dentro del rango, la media es coincidente con el puntaje dado por los docentes en su corrección
- Dentro del rango, desviación coincidente con la calificación dada por los docentes
- 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.

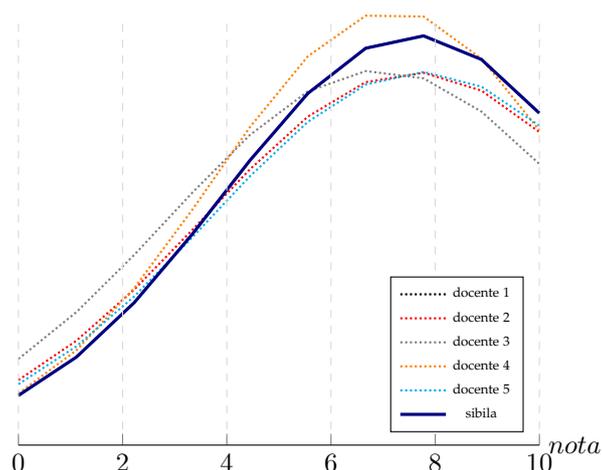


Figura 12: Correlación entre Sibila y los docentes de muestra

6. CONCLUSIONES Y TRABAJOS FUTUROS

Se ha presentado un método 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.

El método presentado se ha implementado en la forma de prototipo de un sistema de corrección automatizado, el que incluye todos los mecanismos descritos en el presente trabajo. En base a los resultados previamente expuestos se ha demostrado que el prototipo ha podido evaluar un conjunto de respuestas en un nivel similar a como lo haría un docente.

El método 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.

Habiendo comprobado que es posible realizar la corrección de respuestas de examen escritas en forma textual, de forma automática, se continuará 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[14].

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

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.

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.

REFERENCIAS

- [1] Marcelo Marciszack y col. *Material de estudio de la Cátedra Paradigmas de Programación*. Ed. por Cátedra de Paradigmas de Programación Ingeniería en Sistemas de Información. Universidad Tecnológica Nacional, Facultad Regional Córdoba, 2016.
- [2] John F Sowa. "Conceptual graphs for a data base interface". En: *IBM Journal of Research and Development* 20.4 (1976), págs. 336-357.
- [3] W Hartshorne. *Burks, editors. Collected Papers of Charles Sanders Peirce, Cambridge, Massachusetts, 1931-1935*.
- [4] Frank Van Harmelen, Vladimir Lifschitz y Bruce Porter. *Handbook of knowledge representation*. Vol. 1. Elsevier, 2008.
- [5] Ivan Olmos, Jesus A Gonzalez y Mauricio Osorio. "Inexact Graph Matching: A Case of Study." En: *FLAIRS Conference*. 2006, págs. 586-591.
- [6] J Glenn Brookshear. *Theory of computation: formal languages, automata, and complexity*. Benjamin-Cummings Publishing Co., Inc., 1989.
- [7] 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).
- [8] 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), págs. 130-140.
- [9] John F Sowa. "Semantic Networks". En: *Encyclopedia of Cognitive Science*. John Wiley & Sons, Ltd, 2006. ISBN: 9780470018866. DOI: 10.1002/0470018860.s00065. URL: <http://dx.doi.org/10.1002/0470018860.s00065>.
- [10] John F. Sowa. "Conceptual Structures". En: ed. por Timothy E. Nagle y col. Upper Saddle River, NJ, USA: Ellis Horwood, 1992. Cap. Conceptual Graphs Summary, págs. 3-66. ISBN: 0-13-175878-0. URL: <http://dl.acm.org/citation.cfm?id=168857.168864>.

- [11] HK Tabbers y col. "Interface design for digital courses". En: *Integrated E-learning: Implications for pedagogy, technology and organization* (2004), págs. 100-111.
- [12] 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), págs. 216-227.
- [13] Fred Buckley y Frank Harary. *Distance in graphs*. Addison-Wesley Longman, 1990.
- [14] Gloria Patricia, Sandra Cristina y col. "Propuesta para la medición del impacto de las TIC en la enseñanza universitaria". En: (2011).