

ANÁLISIS DE PATRONES EN UN GRAFO CONCEPTUAL CONSTRUIDO A PARTIR DE RESPUESTAS ESCRITAS EN FORMA TEXTUAL A PREGUNTAS SOBRE UN TEMA ESPECÍFICO.

María Alejandra Paz Menvielle, Universidad Tecnológica Nacional. F.R.C,
pazmalejandra@gmail.com

Cynthia Lorena Corso, Universidad Tecnológica Nacional. F.R.C,
corso.cynthia@gmail.com

Martín Gustavo Casatti, Universidad Tecnológica Nacional. F.R.C, mcasatti@gmail.com

Analía Guzmán, Universidad Tecnológica Nacional. F.R.C, analia.guzman@the-
group.com.ar

Karina Ligorria, Universidad Tecnológica Nacional. F.R.C, karinaligorria@hotmail.com

Resumen— El objetivo de este trabajo es el análisis de patrones a partir de grafos dirigidos, tanto en sus aspectos teóricos como prácticos y en sus aplicaciones, de tal manera que sirva para la detección de patrones asociados en una base de conocimiento representada en una base de datos orientada a grafos. La misma contiene preguntas y respuestas de exámenes relacionados a los contenidos de la cátedra de Paradigmas de Programación, perteneciente al segundo año de la carrera de Ingeniería en Sistemas de Información y dictada en la Facultad Regional Córdoba de la Universidad Tecnológica Nacional. El contenido de la base de conocimiento se generó a partir de las preguntas y sus respuestas relacionadas, escritas en forma textual sobre un dominio específico. Se comenzará con el estudio de los diferentes patrones de grafos que puedan ser relevantes en la búsqueda de información en el dominio elegido, analizando si tienen comportamientos recurrentes o subyacentes. Esto posibilitará descubrir patrones asociados a la evolución de la base de conocimiento que se utilizarán para determinar el nivel de profundidad con el que se evalúa cada tema y si los distintos exámenes son consistentes en alcance y profundidad de evaluación de conceptos. Esto permitirá, en última instancia, mejorar los materiales didácticos empleados como así también ajustar los instrumentos de evaluación.

Palabras clave— *grafos – patrones – rutas – evaluación.*

1. Introducción

El presente trabajo forma parte del proyecto de investigación y desarrollo que ha sido homologado por la Secretaría de Investigación, Desarrollo y Posgrado de la Universidad Tecnológica Nacional, reconocido con el código UTN4812, el mismo forma parte del Centro de Investigaciones, Desarrollo y Transferencia de Sistemas de Información – CIDS.

Análisis de patrones en un grafo conceptual construido a partir de respuestas escritas en forma textual a preguntas sobre un tema específico.

Para su desarrollo se utiliza como caso testigo a la cátedra de Paradigmas de Programación, perteneciente a la carrera Ingeniería en Sistemas de Información, dictada en la Facultad Regional Córdoba, de la Universidad Tecnológica Nacional.

Para realizar el análisis de patrones a partir de grafos dirigidos, se cuenta con una base de conocimiento modelada como grafo, que posee preguntas y respuestas de exámenes escritas en forma textual, así como cualquier otro concepto contenido en el programa analítico de la materia Paradigmas de Programación.

El primer grafo, generado por los docentes de la cátedra, posee los conceptos del programa, las preguntas y las respuestas de los exámenes y es utilizado para calificar a los alumnos. Es una base de conocimientos amplia y extensible que se utilizará como fuente para la búsqueda de patrones relevantes relacionados a los contenidos del programa analítico.

El segundo grafo, obtenido a partir de las respuestas de todos los alumnos que realizan los exámenes, almacena el historial de las respuestas de los alumnos.

Ambos grafos serán cotejados de tal forma que el análisis de los patrones posibilite descubrir algunas características importantes que se relacionan con las respuestas de los alumnos y que puedan reflejar la evaluación y el aprendizaje de los mismos.

Con el análisis de patrones se pretende, en general, que se pueda determinar el nivel de profundidad con el que se evalúa cada tema, si los distintos exámenes son consistentes en alcance y profundidad de evaluación de conceptos, entre otros. En relación a los contenidos de la materia, se pretende identificar si hay áreas de conocimiento evaluadas con mayor frecuencia, si hay temas que no se están evaluando, qué conceptos de la materia son mejor conocidos por los alumnos y cuáles lo son en menor medida. Esto permitirá, en última instancia, mejorar los materiales didácticos empleados como así también ajustar los instrumentos de evaluación.

2. Materiales y Métodos

2.1 Patrones y grafos

Un patrón es una entidad a la que se le puede dar un nombre y que está representada por un conjunto de propiedades medidas y las relaciones entre ellas (vector de características) [1].

En el dominio utilizado como caso testigo, por ejemplo, un patrón puede ser la ruta resultante de una respuesta de un alumno, de las cuales se extrae el vector de características formado por un conjunto de valores numéricos que pueden representar nivel de exactitud de la respuesta, la puntuación de la misma, la cantidad utilizada de conceptos y de relaciones, etc.

El reconocimiento automático, descripción, clasificación y agrupamiento de patrones son actividades importantes en una gran variedad de disciplinas científicas, como biología, psicología, medicina, visión por computador, inteligencia artificial, teledetección, etc.

Análisis de patrones en un grafo conceptual construido a partir de respuestas escritas en forma textual a preguntas sobre un tema específico.

Lo importante de detectar patrones en los datos es que se pueden inferir causas para la agrupación de los mismos (en el caso de que estemos detectando patrones ya conocidos y ya estudiados).

Este trabajo, tiene paralelos con el llamado SNA (Social Network Analysis, Análisis de Redes Sociales) que es una disciplina cuyo objetivo es “Analizar la estructura de una red social para ‘inferir conocimiento’ de un individuo, un grupo, o las relaciones entre ellos” [2].

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

Se describen a continuación conceptos fundamentales relacionados a la construcción de grafos:

Nodo: Un nodo es un ‘ítem de conocimiento’ (término, entidad) diferente de cualquier otro conocimiento en el modelo. Todos los conceptos que tienen su propio significado están representados como nodos. Un nodo representa la mínima cantidad de conocimiento que puede ser representada, la cual no puede ser dividida.

Unidades de conocimiento mayores pueden representarse mediante grupos de nodos relacionados. Es importante mencionar que los nodos no representan grupos de entidades.

Para los fines que se persiguen es necesario que cada nodo pueda ser identificado de manera unívoca y diferente de todos los demás nodos de la red.

Enlace: Un enlace se utiliza para establecer una relación entre dos nodos del modelo.

Un enlace solo puede conectar dos nodos. Uno denominado ‘origen’ desde el que sale el enlace y uno denominado ‘destino’ al cual llega. Una relación entre un nodo de origen y dos nodos de destino requiere de dos enlaces, ambos con el mismo origen, pero con diferentes destinos. Un nodo puede ser origen de varios enlaces, así como puede ser destino de varios enlaces [5].

2.2 Patrones en Grafos

El reconocimiento o detección de patrones dentro de grafos busca detectar un subgrafo (patrón) en un grafo (objetivo). Debemos considerar que esta búsqueda de coincidencias se puede descomponer en dos partes:

1. Una concordancia estructural, en donde los nodos y relaciones del patrón conforman una estructura existente en el grafo objetivo
2. Una concordancia a nivel de elementos, en donde los nodos y relaciones, a nivel de sus atributos particulares, tiene los mismos valores que en la estructura encontrada en el grafo objetivo.

Muchas veces la búsqueda de estas dos concordancias se ejecuta de forma separada para optimizar los algoritmos o reducir el espacio de búsqueda [6].

En el dominio bajo estudio, la detección de un subgrafo (patrón), se realizará en los grafos (objetivos) que representan los contenidos de la materia y las respuestas de los alumnos en las instancias de exámenes.

2.3. Métricas en grafos

Una herramienta ampliamente utilizada para describir grafos y que muchas veces se utiliza para iniciar el análisis de patrones existentes en los mismos, es el cálculo de métricas [7], locales o globales, que permiten caracterizar el grafo objetivo o el grafo patrón. Las métricas se pueden dividir en dos grandes grupos:

- **Métricas estáticas:** Cuando se calculan sobre un grafo estático en un punto en el tiempo determinado. Se enfocan principalmente en las características estructurales del mismo.
- **Métricas dinámicas:** Tienen en cuenta la dimensión temporal de los cambios que se producen sobre el grafo. Están más enfocadas en las variaciones entre dos instantes de tiempo, antes que en las características propias del grafo en cada uno de esos instantes.

Otro enfoque para el análisis de las métricas radica en analizar sobre qué componentes del grafo se realizan las mediciones. Desde este punto de vista se tienen diversas perspectivas, siendo las más comunes:

- **Métricas de redes (o globales):** Son las métricas que toman como referencia el grafo completo, con todos los nodos y arcos que lo conforman.
- **Métricas nodos (o locales):** Son aquellas que toman como referencia un nodo o subconjunto de nodos para realizar los cálculos.

A continuación, se detallan algunas métricas más comunes:

Métricas globales:

Centralidad: Esta métrica trata de determinar que nodo o nodos ocupan una ubicación central en la red, estando equidistante de los demás nodos.

Conexionado: Busca establecer el grado en el que los nodos de un grafo están conectados con todos los demás nodos del mismo. Se puede encontrar, aplicando esta métrica, componentes fuertemente conectados o débilmente conectados.

Cantidad de Componentes: En un grafo que no es completamente conexo, indica la cantidad de subgrafos conexos que forman parte del grafo. Un componente es un conjunto de nodos conectados que forman parte del grafo principal.

Tamaño del componente gigante: Mide la cantidad de nodos que tiene el componente conectado que es mayor que todos los demás componentes del grafo. En un grafo conexo el tamaño del componente gigante es igual a la cantidad total de nodos.

Ruta más corta/larga: Expresa la longitud (en arcos) mínima/máxima entre dos nodos dados.

Métricas locales:

Conectividad: Expresa la cantidad de conexiones que posee un nodo determinado. Se puede expresar como 'grado', si no tiene en cuenta la dirección de los arcos que inciden o salen del nodo, o como 'grado de entrada' o 'grado de salida' cuando solamente tiene en cuenta los arcos entrantes o salientes, respectivamente.

Centralidad: Es una métrica, asociada a un nodo en un grafo, que determina su importancia relativa dentro de éste, pudiendo dividirse en:

- **Centralidad de grado:** Cantidad de conexiones con otros nodos

Análisis de patrones en un grafo conceptual construido a partir de respuestas escritas en forma textual a preguntas sobre un tema específico.

- Centralidad de cercanía: Indica qué tan cerca se encuentra una unidad de la red de otras.
- Centralidad de intermediación: Indica si una unidad se encuentra dentro de algunas de las rutas más cortas que existen entre dos nodos de la red.

2.4 Análisis de Patrones

El análisis de patrones en el dominio bajo estudio, puede determinar si existen ciertos patrones que, aún, no siendo comunes en otras áreas de la teoría de grafos, si lo son recurrentes en este dominio. Se pueden determinar si son patrones temporales, es decir que tiendan a desaparecer en el tiempo a medida que la base de conocimientos va cambiando, o si son patrones permanentes y/o que se van reforzado con el tiempo.

Dicho análisis puede servir para descubrir algunas características importantes que se relacionan con el aprendizaje, entre ellas:

- los temas que revisten más dificultad de aprendizaje,
- la cantidad y tipos de errores más comunes y su relación con el tema o concepto evaluado,
- las tendencias de los alumnos al momento de responder las mismas preguntas, es decir, si lo hacen con los mismos conceptos o, por el contrario, tienen una riqueza expresiva alta.
- Se propone incluir una respuesta textual, utilizada como patrón, para poder determinar si las respuestas dadas por los alumnos tienen una correspondencia directa (literal) con respecto al material brindado para su estudio.

3. Resultados y Discusión

El diseño de un sistema de reconocimiento de patrones se lleva a cabo normalmente en tres fases [8]:

1. Adquisición y preproceso de datos.
2. Extracción de características.
3. Toma de decisiones o agrupamiento.

3.1 Adquisición y preproceso de datos.

En la introducción del presente trabajo, se mencionó que se cuenta con la base de conocimiento de la materia, donde están almacenados en un grafo, los datos adquiridos y procesados relacionados a los contenidos, las preguntas y las respuestas, tanto las provistas por los docentes como las respondidas por los alumnos.

Esta infraestructura surge de un proyecto anterior, según se ven reflejados en algunos trabajos publicados [9] [10], donde se exponen la arquitectura, las tecnologías utilizadas y los resultados obtenidos a través de una evaluación realizada por medio de un prototipo desarrollado para registrar los datos en una base de datos de grafos y realizar las consultas necesarias para la calificación.

Análisis de patrones en un grafo conceptual construido a partir de respuestas escritas en forma textual a preguntas sobre un tema específico.

La tecnología que se utilizó para crear el prototipo incluye herramientas de código abierto, ya que el uso principal se realiza en un ambiente universitario y por eso es de fundamental importancia no depender de ningún tipo de licenciamiento propietario.

La implementación se realizó en el lenguaje Java, ya que este posibilita implementar la arquitectura en multicapa e intercambiar fácilmente los 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.

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.

Se utilizó, además, la librería gráfica GraphStream, para realizar la visualización de la base de grafos, 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.

Para tener información relevante del dominio elegido y poder detectar los patrones analizados, se agregó a la base de grafos original, los datos de los contenidos de la materia y, además, se agregó una base de grafos, que contiene las respuestas de los alumnos y que se registra automáticamente cuando el sistema califica las respuestas de los alumnos en las distintas instancias de los exámenes.

3.2 Extracción de características y agrupamiento

Para las dos siguientes fases, extracción de características y toma de decisiones o agrupamiento, se consideraron los objetivos y contenidos de la materia, las necesidades de los docentes de la cátedra, respecto a la información que les podría brindar el prototipo de calificación y en general, las características que son de interés para evaluar o estudiar los mecanismos de evaluación de la materia y el grado de aprendizaje de los alumnos.

También se consideraron algunas métricas mencionadas en el apartado anterior, relacionadas a los patrones inicialmente identificados.

Se identificaron dos grandes clases de patrones, aquellos relacionados a la evaluación y aquellos relacionados al aprendizaje de los alumnos.

A continuación, se detalla cada clase de patrones, junto con sus patrones asociados.

Patrones asociados con la evaluación

Estos patrones buscan caracterizar los mecanismos de evaluación utilizados por los docentes y establecer criterios para su mejora tanto en cuanto a profundidad de los temas evaluados como a variedad y relevancia de los conceptos utilizados en dicha evaluación.

- **Conceptos centrales:** Se consideran conceptos centrales a aquellos que tengan más relaciones que la media con conceptos relacionados. De esta manera se busca identificar a aquellos conceptos que son muy utilizados en el campo de conocimiento y cuya evaluación debe ser precisa y valiosa.
- **Conectores:** Se trata de conceptos que, si bien pueden tener pocas relaciones, generalmente aparecen como nexos entre grupos de conceptos importantes. Son de particular importancia para analizar las relaciones que existen entre conjuntos de temas o unidades temáticas.
- **Estructuras conceptuales:** Se considera que una estructura conceptual es la mínima unidad de conocimiento evaluable. Está formada por dos conceptos y una relación que los une. El análisis de la cantidad de estructuras conceptuales existente en el grafo puede dar una idea de la cantidad de conocimiento "evaluable" que hay en el mismo.
- **Conceptos autoreferenciados:** Se considera que un concepto es autoreferenciado si siguiendo una determinada cantidad de relaciones se puede volver sobre el mismo. Dichos conceptos deben ser evitados, sobre todo si la longitud del ciclo que vuelve sobre el mismo es muy corta ya que limita la posibilidad de que los alumnos se explayen en una respuesta. Se plantea en el presente proyecto establecer un tamaño de ciclo mínimo y detectar todos aquellos conceptos que se autoreferencien por medio de una cantidad de relaciones menor a la indicada.
- **Agujeros negros:** Se definen de esa manera a las zonas del grafo que contienen conceptos o estructuras que no forman parte de las respuestas provistas por los docentes o que aparecen en una pequeña cantidad de las mismas, indicando contenidos que no son evaluados o lo son con poca frecuencia. La detección de este patrón es de suma importancia para lograr una evaluación equilibrada de todos los contenidos de la materia.

Patrones asociados con el aprendizaje

Los patrones asociados al aprendizaje tienen como finalidad analizar los contenidos asimilados por los alumnos, traducidos en las respuestas provistas a las preguntas de evaluación, detectar los conceptos recurrentes, los errores más comunes, los temas con dificultades de aprendizaje, las estructuras comunes existentes en las respuestas, entre otros. Se pretende con esto detectar problemas de aprendizaje que posibiliten mejorar tanto el dictado de la materia como la construcción de los instrumentos de evaluación.

Entre los patrones que se buscan se encuentran aquellos caracterizados de la siguiente forma:

Análisis de patrones en un grafo conceptual construido a partir de respuestas escritas en forma textual a preguntas sobre un tema específico.

- **Los temas que revisten mayor dificultad de aprendizaje:** Aquellas estructuras de respuesta que tienen menor puntaje en la evaluación. Se puede analizar en este caso la respuesta total o ciertos conceptos particulares que pueden ser confusos para los alumnos.
- **Cantidad y tipo de los errores más comunes:** Relacionada muy cercanamente con el patrón anterior, busca determinar de qué manera los alumnos suelen equivocarse y sobre qué conceptos, por ejemplo, utilizando sinónimos que no son válidos, ejemplificando de manera errónea, o estableciendo relaciones inválidas entre dos conceptos válidos.
- **Tendencias de los alumnos con respecto a las mismas preguntas:** A través del análisis de diversas respuestas a una misma pregunta base, se busca determinar si los alumnos utilizan de forma uniforme siempre los mismos conceptos o si son capaces de utilizar conceptos relacionados, sinónimos y otras estructuras, para lograr una riqueza conceptual y una variedad expresiva alta.
- **Comparativa con definiciones del material de estudio:** Se propone establecer una respuesta base exacta, que replique exactamente las definiciones existentes en el material de estudio, para verificar si ante la pregunta docente, el alumno elabora una respuesta o simplemente transcribe la definición aprendida de memoria.

4. Conclusiones y recomendaciones

El reconocimiento de patrones es un ámbito de gran auge en la actualidad, si bien la mayoría de las aplicaciones se orientan al reconocimiento de patrones sobre imágenes u otro tipo de información bidimensional, el reconocimiento sobre grafos está cobrando cada vez mayor importancia debido en gran medida al auge de las redes sociales, naturalmente modeladas como grafos, principalmente en la búsqueda de información valiosa que ya existe en la estructura de datos y que se pueda aprovechar.

En el caso particular de este proyecto de investigación se busca detectar patrones subyacentes tanto en la información del programa de la materia como de los exámenes elaborados por los docentes o en los conocimientos apprehendidos por los estudiantes, de forma tal que se pueda mejorar el dictado de la materia y mejorar el aprendizaje, objetivo este central en cualquier proceso educativo.

Se estima que los patrones previamente mencionados brindarán a la cátedra de Paradigmas de Programación un sustento fáctico sobre el cual basar las posibles modificaciones a los materiales de estudio a los métodos de dictado y a los instrumentos de evaluación, así como una herramienta tecnológica eficaz para evaluar, en un tiempo relativamente breve, el impacto de esas modificaciones sobre el cursado y la evaluación.

Se proseguirá la investigación de forma tal que los docentes cuenten con herramientas informáticas para realizar consultas sobre la base de datos de grafos, que ayuden a detectar patrones diferentes a los actualmente planteados. El objetivo final es que los propios docentes puedan indicarle al sistema, a modo de aprendizaje supervisado, qué se debe buscar de acuerdo a los objetivos de la cátedra y tener una realimentación

Análisis de patrones en un grafo conceptual construido a partir de respuestas escritas en forma textual a preguntas sobre un tema específico.

inmediata que posibilite el análisis de diversos escenarios a la hora de plantear herramientas de evaluación y planificar el dictado.

5. Referencias

- [1] WATANABE S. (1985). *Pattern Recognition: Human and Mechanical*. Wiley, New York.
- [2] SCOTT, JOHN; CARRINGTON, PETER J. (2011). *The SAGE handbook of social network analysis*. SAGE publications.
- [3] SOWA, JOHN F. (1992). *Conceptual graph summary*, *Conceptual Structures: Current Research and Practice*. Ellis Horwood, New York London Toront, pp. 3-66.
- [4] PEIRCE, CHARLES S. (1909). *Existential graphs*, *Collected Papers of Charles Sanders Peirce* 4, pp. 1-7.
- [5] PAVLIĆ, MILE; MEŠTROVIĆ, ANA; JAKUPOVIĆ, ALEN. (2013). *Graph-based formalisms for knowledge representation*. *Proceedings of the 17th world multi-conference on systemics cybernetics and informatics (WMSCI 2013)*. p. 200-204.
- [6] FAN, WENFEI. (2012). *Graph pattern matching revised for social network analysis*. *Proceedings of the 15th International Conference on Database Theory*. ACM, p. 8-21.
- [7] VAN STEEN, MAARTEN. (2010). *An Introduction to Graph Theory and Complex Networks*. Copyrighted material.
- [8] ALONSO ROMERO, LUIS; CALONGE CANO, TEODORO. (2001). *Redes neuronales y reconocimiento de patrones*.
- [9] MARIA ALEJANDRA PAZ MENVIELLE Y COL. (2016). *Arquitectura y operatoria de un sistema de corrección de exámenes automatizado, utilizando grafos dirigidos*. IV Congreso Nacional de Ingeniería Informática y Sistemas de Información, CONAIISI, Universidad Católica de Salta, Facultad de Ingeniería, Argentina.
- [10] MARIA ALEJANDRA PAZ MENVIELLE Y COL. (2017). *Teoría y práctica de un sistema de corrección automatizada, utilizando grafos dirigidos como base de conocimiento*. V Congreso Nacional de Ingeniería Informática y Sistemas de Información, CONAIISI, Universidad Tecnológica Nacional Facultad Regional Santa Fe, Argentina.