

Análisis y detección de patrones en un grafo conceptual construido a partir de respuestas escritas en forma textual a preguntas sobre un tema específico - Fase 2.

María Alejandra Paz Menvielle, Calixto Maldonado, Karina Ligorria, Analía Guzmán, Martín Casatti, Nicolás Horenstein

Departamento de Ingeniería en Sistemas de Información
CIDS – Centro de Investigación Desarrollo y Transferencia
de Sistemas de Información
Facultad Regional Córdoba – Universidad Tecnológica Nacional
Maestro Marcelo López esq. Cruz Roja Argentina – Córdoba
0351 – 4686385

pazmalejandra@gmail.com, calixtomaldonado@hotmail.com,
karinaligorria@hotmail.com, aguzman@frc.utn.edu.ar, mcasatti@frc.utn.edu.ar,
nicolashorenstein@gmail.com

Resumen

En este trabajo se presenta una línea de Investigación que busca ampliar la detección de patrones asociados a la evolución de una base de conocimiento representada en una base de datos orientada a grafos, la misma contiene respuestas de exámenes en formato de texto de redacción libre relacionadas a un dominio específico, utilizada para realizar el análisis de texto en respuestas a preguntas de exámenes en la cátedra de Paradigmas de Programación, con el objetivo de detectar el grado de acierto de las respuestas de los alumnos. Dentro de los patrones que se pretenden descubrir se encuentran aquellos asociados a las respuestas de los

alumnos, a la forma de representación de las preguntas de los docentes, entre otras. Es por ello que el presente proyecto busca avanzar en la línea de investigación relacionada a la detección de patrones a partir de grafos dirigidos, tanto en sus aspectos teóricos como prácticos y en sus aplicaciones.

Palabras clave: grafos – patrones – métricas - evaluación.

Contexto

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 SIUTICO0007786TC, el mismo forma parte del Centro de Investigaciones, Desarrollo y Transferencia de Sistemas de Información – CIDS.

Para su desarrollo se utilizará 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. Se respetarán los contenidos mínimos fijados para esta asignatura tal cual figuran en la ordenanza 1150 de la carrera, los cuales pertenecen al bloque de tecnologías básicas dentro del área programación, que están principalmente referidos a los paradigmas lógicos, funcional y de orientación a objetos. Además se cumplirá con los descriptores y criterios de intensidad de formación práctica de la Resolución Ministerial 786/09, los que se encuentran definidos en el área de tecnologías básicas, sub-área programación que incluyen a los paradigmas y lenguajes de programación.

El trabajo que aquí se presenta es la fase 2 del proyecto denominado “Análisis y detección de patrones en un grafo conceptual construido a partir de respuestas escritas en forma textual a preguntas sobre un tema específico”, PID SIUTNCO0004812 [1]. El cual tenía como finalidad analizar, detectar y evaluar patrones topológicos frecuentes en un grafo conceptual. Durante el transcurso del mismo se trabajó principalmente en dos etapas, la adquisición y preprocesamiento de datos, y la extracción de características y agrupamiento con el fin de obtener

algunas métricas y patrones relacionados al grafo conceptual obtenido.

El presente proyecto propone extender la base de datos agregando información en nuevos grafos para cotejarlos con el grafo original, a fin de identificar y mostrar una variedad más amplia de patrones para la gestión del conocimiento de la materia Paradigmas de Programación.

Introducción

Análisis de patrones

Existen muchos fenómenos diferentes que pueden ser representados con grafos y debido a que estos pueden tener grandes dimensiones se hace importante poder extraer información implícita de estas representaciones para una mejor comprensión. La minería de grafos [2] se enfoca en encontrar elementos o patrones representativos dentro de grafos. Una de las tareas de la minería de grafo es extraer subestructuras repetitivas o subgrafos frecuentes dentro de grafos [3]. La minería de subgrafos frecuentes es una rama dentro de la minería de patrones frecuentes. La minería de subgrafos frecuentes sobre colecciones de grafos ha demandado el diseño de métodos especializados [4]. Las colecciones utilizadas en los mismos se han denominado bases de grafos o conjuntos de grafos. Mientras que la minería de datos se enfoca principalmente en los valores de los datos que se están buscando, en los esquemas semi-estructurados y en los grafos, el enfoque se encuentra en etiquetas frecuentes y topologías comunes [5].

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) [6]. 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 la biología, la psicología, la medicina, la visión por computador, la inteligencia artificial, la teledetección, etc. 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” [7]. Un grafo conceptual [8] es un sistema de notación simbólica y de representación del conocimiento. Presentado por John F. Sowa, se basa en los gráficos existenciales [9] de Charles Sanders Pierce, en las estructuras de redes semánticas y en datos de la lingüística, la filosofía y la psicología.

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: una concordancia estructural, en donde los nodos y relaciones del patrón conforman una estructura existente en el grafo objetivo; 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 [10]. 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.

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 [11], ya sean locales o globales, que permiten caracterizar el grafo objetivo o el grafo patrón. También las métricas se pueden dividir en dos grandes grupos: Métricas estáticas y Métricas dinámicas

Líneas de Investigación, Desarrollo e Innovación

- Estudio de los diferentes patrones topológicos de grafos que puedan ser relevantes en la búsqueda de información en el dominio elegido, analizando si dichos patrones tienen comportamientos recurrentes o subyacentes.
- Estudio de algoritmos que permitan detectar patrones conocidos en la teoría de grafos como son las “comunidades”, “pares”, “rutas principales” y otros patrones comunes por medio del análisis de las métricas de la base de conocimientos.
- Detección de patrones que, aún no siendo comunes en otras áreas de la teoría de grafos, si lo son recurrentes en el dominio bajo estudio.
- Descripción de algunas características importantes que se relacionan con el aprendizaje a partir de la aplicación de los patrones encontrados sobre las respuestas elaboradas por los alumnos en un examen.
- Estudio de herramientas de visualización y análisis de grafos para realizar los análisis preliminares y la determinación de los parámetros y métricas de la base de datos [9].
- Automatización de algunos de estos análisis para incluirlos en una herramienta ad-hoc.

Resultados y Objetivos

El objetivo del presente estudio es “Detectar patrones subyacentes en la base de conocimientos asociada a la materia Paradigmas de Programación,

relacionados a los contenidos y a las respuestas a preguntas de examen, que permitan caracterizar diversos comportamientos de docentes y alumnos y obtener información relevante para mejorar el dictado y la gestión de dichos contenidos” Para ello se han identificado los siguientes objetivos particulares:

1. Identificar patrones topológicos en diferentes grafos; que permitan brindar información relevante para la gestión del conocimiento de la materia Paradigmas de Programación.
2. Proponer un conjunto de herramientas que permitan ampliar la base de conocimiento; con nuevos grafos, para representar los contenidos de la materia y las respuestas dada por los alumnos en diferentes instancias de evaluación, tanto las respuestas correctas y las incorrectas.
3. Actualizar el software de preprocesamiento de textos existente para mejorar los resultados del análisis ortográfico, sintáctico y semántico, como paso previo a la carga de información textual en el grafo conceptual para su análisis.
4. Analizar los patrones identificados con el fin de elaborar informes relacionados al comportamiento de los mismos en el dominio de aplicación.

Formación de Recursos Humanos

Dentro del desarrollo de este proyecto de investigación se desarrollará el trabajo de Tesis de Maestría de dos integrantes docentes del presente proyecto. Se incorporan al equipo de trabajo docentes-investigadores de la

carrera Ingeniería en Sistemas de Información como investigadores de apoyo con la finalidad de que inicie su formación en investigación científica y tecnológica, se incorpora un becario graduado BINID y becarios alumnos, quienes colaborarán en la recolección, manipulación y desarrollo de este marco metodológico. En el marco del proyecto los estudiantes tendrán la posibilidad de hacer la Práctica Supervisada de quinto año. Los avances, propuestas y herramientas construidas, estarán disponibles para su transferencia y aplicación en el Centro de Investigaciones, Desarrollo y Transferencia de Sistemas de Información - CIDS. Del mismo modo la detección de patrones sobre el dominio de conocimiento de la materia Paradigmas de Programación continuará beneficiando a los integrantes de la cátedra y a los estudiantes.

Referencias

- [1] PAZ MENVIELLE, A. y col. (2018) “Análisis y detección de patrones en un grafo conceptual construido a partir de respuestas escritas en forma textual a preguntas sobre un tema específico”, XX Workshop de Investigaciones en Ciencias de la Computación, Corrientes, Argentina.
- [2] CHAKRABARTI, D., FALOUTSOS (2012), “Graph Mining Laws, Tools and Case Studies”. C. Morgan & Claypool Publishers.
- [3] BERZAL, F. (2007). “Graph Mining”. Department of Computer Science and Artificial Intelligence E.T.S Ingeniería Informática, Universidad de Granada, Granada, España.
- [4] HAN, J., CHENG, H., DONG, X., YAN, X. (2007). “Frequent Pattern Mining: Current Status and Future Directions”. *Data Mining and Knowledge Discovery*, 2007, vol 15, núm. 1, pp. 55-86.
- [5] VANETIK, Natalia; GODES, Ehud; SHIMONY, Solomon Eyal. (2002). “Computing frequent graph patterns from semistructured data. En *Data Mining*”. *ICDM 2003. Proceedings. 2002 IEEE International Conference on. IEEE*, 2002. p. 458-465.
- [6] WATANABE S. (1985). *Pattern Recognition: Human and Mechanical*. Wiley, New York.
- [7] SCOTT, JOHN; CARRINGTON, PETER J. (2011). *The SAGE handbook of social network analysis*. SAGE publications.
- [8] SOWA, JOHN F. (1992). *Conceptual graph summary, Conceptual Structures: Current Research and Practice*. Ellis Horwood, New York London Toront, pp. 3-66.
- [9] PEIRCE, CHARLES S. (1909). *Existential graphs, Collected Papers of Charles Sanders Peirce* 4, pp. 1-7.
- [10] FAN, WENFEI. (2012). *Graph pattern matching revised for social network analysis. Proceedings of the 15th International Conference on Database Theory. ACM*, p. 8-21.
- [11] VAN STEEN, MAARTEN. (2010). *An Introduction to Graph Theory and Complex Networks*. Copyrighted material.