Desarrollo de un agente conversacional para el aprendizaje de SQL

Development of a chatbot for learning SQL

Antonio Balderas

Departamento de Ingenier´ıa Informática, Universidad de Cádiz Puerto Real, Spain antonio.balderas@uca.es

Tatiana Person

Departamento de Ingenier'ia Informática, Universidad de Cádiz Puerto Real, Spain tatiana.person@uca.es Daniel Mej'ıas-Ram'ırez
Departamento de Ingenier'ıa
Informática, Universidad de Cádiz
Puerto Real, Spain
daniel.mejiasramirez@alum.uca.es

Manuel Palomo-Duarte

Departamento de Ingeniería Informática Universidad de Cádiz Puerto Real, Spain manuel.palomo@uca.es

Iván Ruiz-Rube

Departamento de Ingeniería Informática Universidad de Cádiz Puerto Real, Spain ivan.ruiz@uca.es

Abstract—Learning the SQL language is fundamental in the training of computer engineers. However, its learning often presents difficulties for students, especially when much of its study is done autonomously. In this work, a chatbot is developed for learning SQL. The chatbot will support students to perform a set of SQL exercises while providing help and feedback based on their actions. The chatbot has been used by university students a database course. The results indicate that the chatbot is not only useful for the student, but also for the course coordinator because of the amount of information provided by the tool that can be analyzed.

Resumen—SQL es un lenguaje ampliamente usado, siendo uno de los pilares fundamentales en la formación de los ingenieros en informática. Sin embargo, su aprendizaje suele presentar dificultades para los estudiantes, sobre todo cuando gran parte del estudio de la materia se realiza de manera autónoma. En este trabajo, se desarrolla un agente conversacional (chatbot) para el aprendizaje del SQL. El chatbot, permitirá al estudiante realizar una serie de ejercicios SQL mientras le brinda, por un lado, ayuda para realizar los ejercicios y, por otro lado, retroalimentación en base a las acciones del estudiante. El chatbot ha sido probado con estudiantes universitarios de bases de datos y los resultados indican que el chatbot no solo es útil para el estudiante, sino también para el profesorado por la cantidad de información que la aplicación le proporciona.

Index Terms—sql, chatbot, web, agente conversacional, learning analytics

I. Introducción

Gracias a la aplicación de las tecnologías informáticas, el profesorado puede diseñar experiencias de aprendizaje más atractivas y efectivas para sus estudiantes que le ayuden a alcanzar los objetivos de aprendizaje [1], [2]. Además,

Este trabajo ha sido desarrollado en el contexto del proyecto VISAIGLE (TIN2017-85797-R), financiado por la Agencia Estatal de Investigaci ón del Ministerio de Ciencia e Innovación de España.

en contextos en los que lafigura del profesor no est a´ siempre disponible, como puede ocurrir en la educación a distancia o en situaciones de confinamiento como la vivida como consecuencia de la reciente pandemia originada por el COVID-19, la educación puede pasar a depender en mayor grado de la tecnolog´1a y de su disponibilidad [3].

Una de las herramientas tecnológicas que, tras usarse con éxitos en otros entornos, están empezando a incorporarse dentro de las plataformas de aprendizaje electrónico son los agentes conversacionales [4], [5]. Los agentes conversacionales, también conocidos comochatbots, interactúan con los usuarios usando lenguaje natural y son capaces de responder a preguntas de los usuarios, proporcionándoles las respuestas personalizadas [6] e incluso gestionando las emociones en el diálogo con los estudiantes [7]. Grandes empresas han desarrollado varios chatbots tanto para el ámbito industrial como para la investigación: Apple Siri, Microsoft Cortana, Facebook M e IBM Watson.

Este trabajo tiene su origen en el aprendizaje del lenguaje de consulta estructurado SQL (del inglés Structured Query Language). SQL es un lenguaje de dominio espec´ıfico diseñado para administrar y recuperar información de los sistemas de gestión de bases de datos relacionales [8]. La parte del SQL correspondiente a la manipulación de los datos (DML, del inglés Data Manipulation Language), está basado en el álgebra relacional y en el cálculo relacional de tuplas. En consecuencia, el no disponer de un conocimiento consolidado de estos lenguajes puede suponer un serio inconveniente para el estudiante a la hora de aprender este lenguaje en profundidad [9].

Para ayudar al estudiante en el aprendizaje de SQL, este trabajo propone el empleo de un chatbot. El chatbot gu'ia

al estudiante en la resolución de diversas consultas SQL mediante interacciones en lenguaje natural, motivadas por sus preguntas e intentos. El chatbot será utilizado en una asignatura universitaria de bases de datos.

El resto del art´ıculo se organiza como sigue. La segunda sección repasa varias sistemas software de apoyo el aprendizaje de SQL. En la tercera sección, se describe el chatbot desarrollado. La cuarta sección presenta la experiencia llevada a cabo. Finalmente se muestran las conclusiones y el trabajo futuro.

II. INICIATIVAS PARA EL APRENDIZAJE DE SQL

Las bases de datos son uno de los temas fundamentales para todos los estudiantes de Ingeniería Informática. En un informe de las asignaturas de Bases de Datos en los estudios de Grado en Ingeniería Informática en universidades españolas de 2018 [10], se muestra cómo en 38 de los 54 títulos con asignaturas propias de Bases de Datos se imparte diseño conceptual, en 44 modelo relacional, en 30 álgebra y cálculo relacional, y en 40 diseño lógico. En estas asignaturas, se trabaja con el lenguaje SQL disponible en Sistemas de Gestión de Bases de Datos Relacionales (SGBDR) como Oracle, MySQL y SQLServer.

Aunque las consultas más sencillas de SQL suelen ser fáciles de aprender para los estudiantes, aprender SQL a fondo no suele resultar una tarea sencilla para el estudiante si no tiene una buena base de álgebra y de cálculo relacional. La academia ha tratado de mejorar la situación mediante el uso de herramientas interactivas que ayuden al estudiante en el aprendizaje de SQL [11]. Con este objetivo, presentamos a continuación diferentes herramientas gratuitas que podemos encontrar en Internet y que han sido recomendadas por el profesorado y utilizada por estudiantes.

La herramienta de aprendizaje de SQL más veterana de las analizadas es SQL Course [12]. SQL Course es una herramienta interactiva online que ofrece un curso teórico práctico de SQL, proporcionando la teoría básica de las diferentes cláusulas SQL, diferentes ejemplos y un espacio para que el usuario proporcione una solución a los diferentes ejercicios que propone. Aunque utiliza un enfoque muy pedagógico, la experiencia del estudiante con esta web es muy mejorable en dos aspectos. En primer lugar, al realizar los ejercicios cada consulta produce una recarga en la página, creando tiempos de esperas innecesarios. En segundo lugar, no hay interacción con la aplicación a la hora resolver los ejercicios: si una consulta esta bien, se muestran los resultados y, si está mal, se lanza un mensaje de error genérico.

*SQLzoo*es otra herramienta web interactiva para el aprendizaje de SQL [13]. Esta herramienta es similar a SQL Course, aunque mejora en parte la retroalimentación que recibe el estudiante. Una vez que el estudiante propone una solución a un ejercicio propuesto, la aplicación le devuelve el resultado de la ejecución de la consulta y le indica si dicho resultado es o no el esperado.

En l'inea con las herramientas anteriores, encontramos SQL easy [14]. SQL easy es tambi én una herramienta web que

proporciona feedback sobre la corrección o no de las consultas proporcionadas a los ejercicios propuestos. A diferencia de las anteriores, ofrece un diseño más actualizado y amigable, aunque tampoco guía y apoya al estudiante en el desarrollo de las consultas más allá de informarle de si el resultado de la consulta esta´ correcto o no.

Khanacademyse presenta como una plataforma m ás actualizada y completa que las presentadas anteriormente [15]. No solo proporciona una funcionalidad parecida, sino que además ofrece un seguimiento mediante vídeos interactivos que gu´ian al estudiante para probar las diferentes consultas. Aunque la caracter´istica de los v´ideos es interesante, utiliza un enfoque genérico y no adaptado a las necesidades de cada estudiante. Por otro lado, a la hora de resolver ejercicios tampoco aporta más ayuda que la comprobación de si el resultado de la consulta propuesta esta´ esperado o no.

La personalización de las herramientas de aprendizaje a las necesidades de cada estudiante es objetivo de los docentes. SQL-KnoT es una herramienta que genera problemas que los estudiantes deberán resolver mediante consultas SQL. En [16], SQL-Knot se integró con el sistema WebEx y con SQL-Lab para construir un modelo de usuario para cada estudiante, que se utilizó para adaptar algunos de sus componentes a partir del progreso individual.

III. AGENTE CONVERSACIONAL PARA EL APRENDIZAJE DE SOL

A. Descripción de la funcionalidad

La plataforma desarrollada ofrece dos perfiles diferentes de usuarios: profesor y estudiante. Como estudiante, la plataforma permite realizar ejercicios de SQL, clasificados en tres niveles de dificultad, los cuales se irán habilitando conforme el aprendiz vaya completándolos. Para cada ejercicio, la herramienta muestra su enunciado (véase lafigura 1, recuadro superior central) y, sobre una caja de entrada de texto situada justo bajo el recuadro del enunciado, el aprendiz podra´ lanzar las consultas requeridas y observar los resultados producidos. En la parte derecha de la aplicación, se muestra un panel desde donde el usuario podra' interactuar con el chatbot. Este chatbot le proporcionará pistas sobre cómo resolver los ejercicios, as'icomo instrucciones claras y concisas sobre la sintaxis y semántica del lenguaje SQL. La aplicación incorpora un ranking donde aparecen los usuarios que han completado los ejercicios.

Desde el punto de vista del profesor, la herramienta permite añadir nuevos ejercicios, para lo cual se deberá indicar su enunciado, una consulta SQL que permita generar la solución correcta y un conjunto de pistas adicionales de ayuda al aprendiz. Además, la plataforma ofrece una pantalla (véase lafigura 2) desde donde acceder al conjunto de ejercicios y conversaciones mantenidas entre los estudiantes y el chatbot, con elfin de analizar su comportamiento.

B. Desarrollo de la plataforma

Para el desarrollo del chatbot, se ha tenido en cuenta la necesidad de satisfacer una serie de requisitos no funcionales

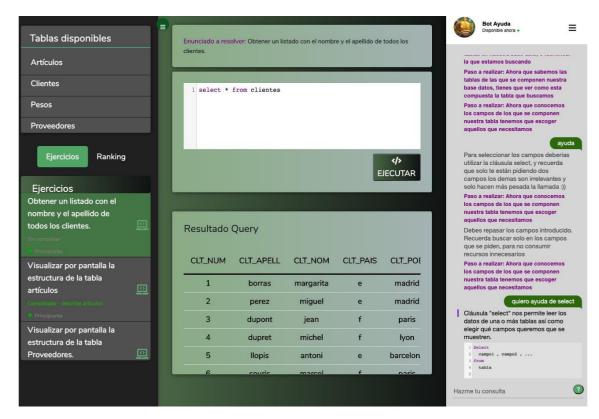


Fig. 1. Pantalla de realización de ejercicios

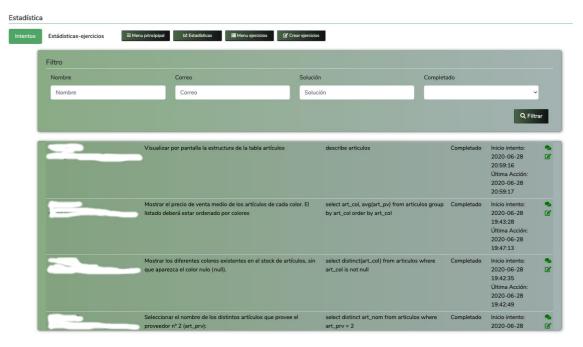


Fig. 2. Pantalla de monitorización de ejercicios

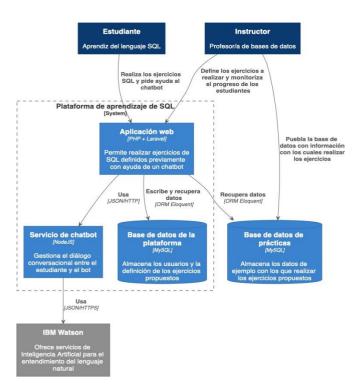


Fig. 3. Arquitectura de la plataforma

que han motivado la elección de la tecnología. En primer lugar, se optó por una aplicación web basada en el framework Laravel, debido a su extensa documentación, claridad y rendimiento en consultas a bases de datos gracias a Eloquent ORM (Object-Relational Mapping) [17].

Para diseñar la inteligencia artificial encargada de proporcionar al chatbot de la capacidad de interpetar el lenguaje natural se ha optado por *IBM Watson*. Se opt ó por *IBM Watson* por sus funciones de monitorización e integración con herramientas de Big Data, además de que es una herramienta cuyo uso está extendido en investigación en educación [18].

Para conectar a los usuarios con la inteligencia artificial de IBM Watson se decidió crear un servidor con*Node.js*, el cual permite que se puedan recibir consultas llamadas de manera concurrente [19].

Finalmente, las bases de datos de trabajo se alojan en *Amazon Web Services*(AWS). La elecci ón de AWS se sustenta en lafiabilidad que proporciona esta plataforma, su facilidad para administrar los servicios contratados y por los aspectos no funcionales que ofrece tan importantes como la seguridad [20].

En lafigura 3, se muestra la arquitectura de la plataforma, en forma de diagrama de contenedores según se propone en el modelo C4 [21].

C. Descripción interacción estudiante-chatbot

La interacción de los estudiantes con el chatbot consiste en una serie de pasos que se describen a continuación:

- Ejercicio. La aplicaci ón le proporciona al alumno el enunciado del ejercicio a resolver.
- 2) **Consejo**. El chatbot proporciona un consejo al estudiante para dar el siguiente paso en el ejercicio.
- 3)**Proponer soluci ón o pedir ayuda**. El alumno puede introducir una consulta para solucionar el ejercicio o pedir ayuda al chatbot (en caso de pedir ayuda, se volvería al paso 2).
- Retroalimentaci ón. Tras introducir la consulta, esta es evaluada por el SGBD y se muestra el resultado devuelto.
 - Si el resultado de la consulta es correcto, el ejercicio se da por resuelto y el estudiante vuelve al menú principal.
 - En caso contrario, el chatbot informa al alumno de que la consulta que ha definido es incorrecta, le proporciona una nueva pista y se vuelve al paso 3.

IV. EVALUACIÓN

Para llevar a cabo la evaluación de la herramienta, se puso a disposición de los alumnos de la asignatura Bases de Datos, obligatoria en segundo curso del Grado en Ingenier´ia Informática de la Universidad de Cádiz durante parte del curso 2019/2020. El objetivo de esta evaluación preliminar, en la que participaron 59 estudiantes, ha sido inspeccionar el comportamiento del chatbot con estudiantes reales para valorar la calidad y efectividad de sus respuestas. La herramienta sirvió como material de refuerzo para la preparación del examenfinal de la asignatura, especialmente adecuada para este curso debido a la suspensión de clases presenciales por la pandemia del COVID19.

A modo de ilustración se describe, paso a paso, una de las interacciones realizadas por uno de los estudiantes con el chatbot. En lafigura 4 se puede observar la conversaci ón mantenida de uno de los estudiantes con el chatbot durante la realización de uno de los ejercicios disponibles en la plataforma.

- 1)**Ejercicio**. La aplicaci ón le solicita al alumno hallar todos los clientes que viven en Madrid.
- 2) Consejo. El chatbot le sugiere al alumno escribir la consulta necesaria para mostrar las diferentes tablas existentes en la base de datos.
- Pedir ayuda. El alumno escribe la consulta para visualizar las diferentes tablas de la base de datos.
- 4)**Proponer soluci ón**. El alumno propone una solución al ejercicio propuesto.
- 5) **Retroalimentaci ón**. El SGBD evalúa la consulta como incorrecta y se muestra el error obtenido. En este caso, el alumno ha utilizado una columna no existente en la tabla consultada de la base de datos.
- 6) Consulta incorrecta. El chatbot informa al alumno de que la consulta que ha definido es incorrecta, indicándole de que no se encuentra la columna a la que aplicar el filtro.
- 7)**Proponer soluci ón**. El alumno propone otra solución al ejercicio propuesto.

Datos del intento Consultas Conversación show tables Hola, vamos a resolver el siguiente ejercicio: describe clientes Hallar todos los clientes que viven en Madrid select * from clientes where ciudad = 'madrid' Paso a realizar: Escribe la consulta necesaria para mostrar las diferentes tablas en nuestra base dato, e identificar la select * from clientes where clt_pob = 'madrid' Paso a realizar: Ahora que sabemos las tablas de las que Errores se componen nuestra base datos, tienes que ver como esta compuesta la tabla que buscamos SQLSTATE[42S22]: Column not found: 1054 Unknown column 'ciudad' in 'where clause' que se componen nuestra tabla tenemos que escoger (SQL: select * from clientes where ciudad = aquellos que necesitamos 'madrid') No se encuentra columna a la que aplicar el filtro so a realizar: Ahora que conocemos los campos de los que se componen nuestra tabla tenemos que escoger aquellos que necesitamos Correcto lo has terminado crack

Fig. 4. Ejemplo de conversación de un alumno con el chatbot durante la realización de un ejercicio

- 8) **Retroalimentaci ón**. El SGBD evalúa la consulta como correcta y se muestra su resultado. En este caso, se listan las instancias de los clientes existentes en la base de datos que viven en Madrid.
- 9) **Consulta correcta**. El chatbot informa al alumno de que la consulta es correcta, el ejercicio se da por resuelto y el alumno vuelve al menú principal.

Tras haber monitorizado las primeras interacciones los estudiantes con el chatbot, se puede comprobar que la herramienta gu'ia a los estudiantes a resolver los errores que contengan sus soluciones propuestas, permitiendo que vuelvan a intentar definir su solución en un número no limitado de ocasiones. Este enfoque puede ser de gran ayuda para los alumnos durante el estudio autónomo de la asignatura de Bases de Datos, ya que no pueden interactuar de forma directa con el profesor de forma tan fácil y rápida. En la tabla I se recogen algunos de las estadísticas obtenidas de forma automática desde la aplicación. La primera columna se corresponde a los ejercicios. La segunda muestra el nivel de dificultad del ejercicio: Principiante (A), Intermedio (B) y Avanzado (C). La columnaResueltosmuestra el n úmero de estudiantes que lo resolvieron con éxito. La cuarta columna Abandonos muestra el número de estudiantes que abandonaron el ejercicio sin resolverlo. La columna Media Errores indica cu ántos errores comete de media cada estudiante a la hora de resolver la consulta. Finalmente, la columna Media Consultas nos indica cuántas consultas de media utilizan los alumnos antes de dar con la solución.

Cabe destacar que para todos los ejercicios, el número de estudiantes que consiguió resolver cada ejercicio fue muy superior al de aquellos que abandonaron su resolución. Del mismo modo, los datos permiten identificar que´ tipo de ejercicios suponen mayor dificultad al estudiante. Por ejemplo,

Ejercicios	Nivel	Resueltos	Abandonos	Errores	Consultas
E1	A	49	7	0,73	3,17
E2	A	54	2	0,34	2,37
E3	В	51	2	0,57	2,83
E4	В	51	2	0,09	1,53
E5	В	52	1	0,34	1,82
E6	В	42	11	1,09	5,00
E7	В	50	3	0,45	3,06
E8	С	36	15	1,52	4,21
E9	С	36	15	1,99	5,74
E10	C	47	4	0,49	2,73

ESTADÍSTICAS DE LOS INTENTOS REALIZADOS DE LOS ESTUDIANTES EN CADA EJERCICIO

vemos que el ejercicio número 9 es el que menos estudiantes han logrado resolver (36), el que más abandono ha tenido (15), en el que los estudiantes tienen una media superior de errores (1,99) y el que requiere a los estudiantes que lo logran resolver un mayor número de consultas.

V. CONCLUSIONES

Aprender el lenguaje SQL es una de las actividades que, con mayor frecuencia, presenta más dificultades para los estudiantes de Ingeniería Informática. Por ello, en este trabajo se presenta la primera versión de un agente conversacional, incrustado en una aplicación web, con el objetivo de ayudar a los aprendices en bases de datos a realizar correctamente consultas SQL. Los resultados, hasta el momento, son prometedores por el alto número de alumnos que, habiendo utilizado el software, han conseguido completar todos los ejercicios propuestos.

En la actualidad, la herramienta no permite lanzar consultas SQL que involucren varias tablas, ya sean para realizar producots (*joins*) o subconsultas. De igual manera, no

reconoce las instrucciones SQL de manipulación de datos *insert*, updatey delete. Como trabajo futuro, se pretende ampliar las capacidades del chatbot para poder reconocer los tipos de sentencias SQL anteriores y permitir la inclusión de nuevos or ígenes de datos para utilizar bases de datos provistas específicamente por cada profesor. Además, se plantea realizar una evaluación más exhaustiva de la herramienta, mediante la realización de un cuestionario de aceptación de tecnología con los estudiantes de Ingeniería Informática de la Universidad de Cádiz y, si fuera posible, con estudiantes de otras universidades.

Finalmente, el uso masivo del chatbot nos permitira´ también aplicar técnicas de aprendizaje automático (*machine learning*)para mejorar diversos aspectos de los procesos de aprendizaje [22]. Gracias a los algoritmos de aprendizaje automático que incorpora IBM Watson y los ajustes que irán realizando los profesores, se irá mejorando, de manera progresiva, la precisión en las respuestas y consejos ofrecidos por el chatbot.

REFERENCES

- A. Balderas, A. Berns, M. Palomo-Duarte, J. M. Dodero, and I. Ruiz-Rube, "Retrieving objective indicators from student logs in virtual worlds," *Journal of Information Technology Research (JITR)*, vol. 10, no. 3, pp. 69–83, 2017.
- [2] E. Labrador, E. Villegas, R. S. Contreras, X. Canaletaet al., "Teaching teamwork in logistics engineering through a board game," The International journal of engineering education, vol. 36, no. 1, pp. 510–520, 2020.
- [3] F. J. García-Peñalvo, V. Abella-García, A. Corell, and M. Grande, "La evaluación online en la educación superior en tiempos de la covid-19," Education in the Knowledge Society, vol. 21, no. 12, 2020.
- [4] S. Tamayo and D. Pérez-Marín, "¿qué esperan los docentes de los agentes conversacionales pedagógicos?" Teor ía de la Educación. Educación y Cultura en la Sociedad de la Información, vol. 18, no. 3, pp. 59–85, 2017.
- [5] G. D'Aniello, A. Gaeta, M. Gaeta, and S. Tomasiello, "Self-regulated learning with approximate reasoning and situation awareness," *Journal* of Ambient Intelligence and Humanized Computing, vol. 9, no. 1, pp. 151–164, 2018
- [6] B. A. Shawar and E. Atwell, "Using dialogue corpora to train a chatbot," in Proceedings of the Corpus Linguistics 2003 conference, 2003, pp. 681–690.
- [7] E. K. Morales-Urrutia, J. M. Ocaña, and D. Pérez-Marín, "How to integrate emotions in dialogues with pedagogic conversational agents to teach programming to children," in *Innovative Perspectives on Interactive Communication Systems and Technologies*. IGI Global, 2020, pp. 66–91.
- [8] A. Beaulieu, Learning SQL: master SQL fundamentals. "O'Reilly Media, Inc.", 2009.
- [9] J. V. Murillo, S. B. Chavarr'1a, and S. M. Rivera, "Herramienta asistida por computadora para la enseñanza del álgebra relacional en bases de datos," *Uniciencia*, vol. 26, no. 1-2, pp. 179–195, 2012.
- [10] F. Carrillo Chaves, "Informe asignaturas de bases de datos en estudios de grado en ingeniería informática en universidades españolas, 2018," 2018.
- [11] C. R. Jaimez-González and A. Palma-Hernández, "An interactive online training course for sql beginnersbeginners," *International Journal on Integrating Technology in Education (IJITE)*, vol. 6, no. 4, pp. 1–9, 2017.
- [12] sqlcourse.com Interactive Online SQL Training, (accessed July 8, 2020). [Online]. Available: http://www.sqlcourse.com/
- [13] A. Cumming, sqlcourse.com Interactive Online SQL Training, 2009 (accessed July 8, 2020). [Online]. Available: https://sqlzoo.net/
- [14] SQLeasy, Easy Interactive Online SQL Training for Beginners, (accessed July 8, 2020). [Online]. Available: https://www.sql-easy.com/

- [15] K. Academy, Unidad: Introducci ón a SQL: consulta y gestión de los datos, (accessed July 8, 2020). [Online]. Available: https://es.khanacademy.org/computing/computer-programming/sql
- [16] P. Brusilovsky, S. Sosnovsky, M. V. Yudelson, D. H. Lee, V. Zadorozhny, and X. Zhou, "Learning sql programming with interactive tools: From integration to personalization," ACM Transactions on Computing Education (TOCE), vol. 9, no. 4, pp. 1–15, 2010.
- [17] M. Stauffer, Laravel: Up & Running: A Framework for Building Modern PHP Apps. O'Reilly Media, 2019.
- [18] S. S. Murtaza, P. Lak, A. Bener, and A. Pischdotchian, "How to effectively train ibm watson: Classroom experience," in 2016 49th Hawaii International Conference on System Sciences (HICSS). IEEE, 2016, pp. 1663–1670.
- [19] S. Tilkov and S. Vinoski, "Node. js: Using javascript to buildhighperformance network programs," *IEEE Internet Computing*, vol. 14, no. 6, pp. 80–83, 2010.
- [20] F. Bracci, A. Corradi, and L. Foschini, "Database security management for healthcare saas in the amazon aws cloud," in 2012 IEEE Symposium on Computers and Communications (ISCC). IEEE, 2012, pp. 000 812–000 819.
- [21] S. Brown, The art of visualising software architecture: communicating software architecture with sketches, diagrams and the C4 model. Lean Publishing, 2016.
- [22] A. Khakpour and R. Colomo-Palacios, "Convergence of gamification and machine learning: A systematic literature," *Technology, Knowledge* and Learning, 2020.