

# Creación automática de enunciados para prácticas de auditoría de protección de datos personales en una asignatura de aspectos legales de la informática

Cristian González García, Darío Álvarez Gutiérrez  
Departamento de Informática  
Universidad de Oviedo  
Asturias, España  
gonzalezcristian@uniovi.es, darioa@uniovi.es

## Resumen

La creación de ejercicios personalizados para cada alumno o grupo de alumnos es una tarea que lleva mucho tiempo al docente (alrededor de un 20%, según algunos estudios). Además, es una tarea repetitiva y relativamente compleja, al requerirse ejercicios de dificultad similar para todos los alumnos. Este era el caso de una asignatura que incluye aspectos legales de la informática, donde para una práctica de auditoría de protección de datos personales se necesitaba crear manualmente los enunciados individualizados para cada grupo de prácticas, definiendo en esencia una empresa ficticia sobre la que realizar la auditoría. Por otro lado, en esta creación manual surgían diferencias de detalle en los enunciados que obligaban a dedicar tiempo adicional a aclarar esos extremos a los alumnos. Por ello se ha desarrollado una herramienta, llamada TextReportGenerator, que permite generar de manera automática enunciados que describen una empresa ficticia, con la mayoría de los detalles necesarios para desarrollar un ejercicio de auditoría de protección de datos personales de la empresa. Con esta herramienta, se ha conseguido reducir de forma amplia el tiempo dedicado por los profesores a la creación de los enunciados, así como el dedicado por profesores y alumnos a la aclaración y complementación de la información proporcionada.

## Abstract

Creating customized exercises for each student or group of students is a task that takes much time for the lecturer (around 20%, according to different studies). Besides, it is a repetitive and relatively complex task, as it is required that exercises should have approximately the same level of difficulty. That was the case of a course that includes legal aspects of computing, where the wordings of customized exercises for a lab task in personal data privacy auditing were created manually, by defining a fictional company that was to be audited. Apart from that, this

manual generation of wordings introduced differences in details that took additional time in order to clarify them to students. Therefore, we developed the TextReportGenerator tool, which is able to automatically generate wordings that describe a fictional company, including most of the details that are required in order to complete a personal data privacy audit. With this tool, the time employed by lecturers to the generation of exercises has been greatly reduced, as well as the time used by lecturers and students to clarify and complement the information provided in the exercises.

## Palabras clave

Generador de enunciados, aspectos legales, protección de datos personales, auditoría, docencia, prácticas, herramienta, software.

## 1. Introducción y motivación

La finalidad de este proyecto es la mejora de la asignatura de Aspectos Sociales, Legales, Éticos y Profesionales de la Informática (ASLEPI), de 4º curso del grado de Ingeniería Informática en Ingeniería de Software de la Escuela de Ingeniería Informática de la Universidad de Oviedo. Uno de los objetivos en esta asignatura es el de «capacitar para el desarrollo de una actividad profesional en el campo del peritaje y la auditoría en informática». Así, teniendo en cuenta además que la asignatura trata aspectos legales relacionados con la protección de datos personales, una de las dos prácticas finales que se piden consiste en realizar una auditoría de protección de datos personales a una empresa ficticia. En total, en el curso 2018-2019 cinco profesores participaron en la asignatura, con sus respectivos grupos de laboratorio de casi 16 alumnos de media. En el curso 2019-2020 los profesores fueron 6.

Para esta práctica de auditoría, el profesor de cada grupo debe crear múltiples enunciados del ejercicio (uno para cada grupo de 3 alumnos): la descripción

de los aspectos de protección de datos personales de una empresa ficticia. Luego, deben generarse aproximadamente 4-5 descripciones de empresas por grupo de laboratorio. Estas descripciones incluyen a la empresa y su sector, el material, la metodología de trabajo, sus características, el software y hardware que utilizan y cómo lo usan, su seguridad y cómo la cumplen o incumplen, la composición de la empresa, su organización física, etc. Con base en esto, cada grupo tiene que realizar una auditoría de protección de datos personales a la empresa, según las especificaciones marcadas por el Reglamento General de Protección de Datos y la LOPD-GDD [1]. Para ahorrar tiempo, esta descripción se indicaba, por lo general, de palabra, lo que hacía que los alumnos tuvieran que tomar nota y quizás no recogieran todos los datos, o, incluso, debido a errores o lagunas, les faltaran datos. Esto se hacía así porque el trabajo para describir por adelantado un enunciado diferente para cada grupo de alumnos se había estimado excesivo. En su lugar, se había pensado que sería más efectivo, de forma global, el determinar las circunstancias variables de cada enunciado en directo en la sesión de prácticas (a partir de un patrón de elementos principales a incluir según el conocimiento del profesor). Otros profesores derivaban el esfuerzo principal de definición de la empresa en el alumno, y posteriormente revisaban el resultado.

En cualquier caso, puede ocurrir que el enunciado no incluya todos los datos necesarios para realizar la auditoría. Se insiste a los alumnos que deben consultar con los profesores (principalmente por correo) para completar los posibles datos faltantes. En algunos casos, los alumnos no detectan o creen irrelevante la falta de algunos datos, produciendo con ello un trabajo incorrecto. Al otro extremo, algunos alumnos envían al profesor decenas de correos solicitando datos faltantes que en realidad no son necesarios. El resultado son prácticas con muy diferentes grados de completitud, estrés en algunos grupos de alumnos, parones en el trabajo debido a falta de datos, y trabajo extra a los profesores contestando correos.

Por otro lado, la creación de muchas y buenas descripciones de empresas que sean diferentes entre sí es muy difícil, tanto entre diferentes grupos de prácticas como en diferentes años. Es habitual que las descripciones se acaben pareciendo mucho entre sí.

Este aspecto de variabilidad en las descripciones es muy importante, ya que el que cada ejercicio sea diferente está muy relacionada con la calidad de la evaluación [3]. Como se indica, la variabilidad ayuda a que los ejercicios se puedan repetir durante varios cursos, evitando el caso habitual en el que las propuestas de diferentes años cambian muy poco y de forma limitada respecto a una serie de descripciones/tareas estándares. Además, si existe suficiente variabilidad, los estudiantes no pueden compartir las respuestas y también permite que los

estudiantes puedan realizar los mismos ejercicios con los que se les evaluará, es decir, se pueden generar tanto ejercicios como exámenes. Así pues, como se trata en [7], la generación de este tipo de ejercicios con variabilidad en las descripciones ayuda a evitar el plagio e, incluso, a frenar las posibles trampas que puedan cometer los estudiantes mediante la generación de ejercicios similares pero diferentes e individualizados para cada alumno [2].

Otro problema es el que afecta a los profesores, pues se estima que el 20% de horas de docencia se destinan a desarrollar enunciados para evaluaciones [3], convirtiéndose en una tarea intelectual repetitiva y rutinaria que requiere gran esfuerzo. La creación de deberes, ejercicios y exámenes lleva tiempo y entraña una cierta dificultad [2]. Estos estudios coinciden con la percepción del esfuerzo dedicado a la generación de enunciados en el caso particular de los profesores de la asignatura objeto de este trabajo. La generación automática de este tipo de enunciados permitiría aliviar significativamente este problema.

Respecto al tema de generación de ejercicios de forma automática existe mucha literatura. No obstante, el foco suele estar en ejercicios relacionados con las matemáticas o en algoritmos. Es decir, cuestiones que parecen más sencillas de automatizar en comparación con la generación de enunciados en lenguaje natural. Esto se debe a que, como se afirma en [6], la generación de ejercicios de forma automática está bastante limitada dependiendo del dominio. Luego, el caso de esta asignatura, que es generar la descripción de la estructura y funcionamiento de una empresa (con respecto a la protección de datos personales) no es sencillo, máxime cuando se deben generar numerosos enunciados diferentes.

Dentro del campo de la informática, la mayoría de los trabajos están enfocados en la programación, y suelen limitarse al uso de plantillas para generar preguntas parametrizadas [6]. En [8] se cambian palabras y sentencias en problemas de programación para producir diferentes variaciones. En [9], se usan plantillas para generar trazas y ejercicios de depuración para enseñar la gramática y la semántica de C. En [5] se parametrizan programas para enseñar Java. De acuerdo con [6], estas técnicas son útiles para prevenir el plagio, pero no son flexibles para crear ejercicios individuales.

Como se ha visto, la gran mayoría de la literatura relacionada suele generar ejercicios parametrizables. Mientras, algunas técnicas similares que usan texto y que se usan para estudiar idiomas se basan en extraer información de textos y crear huecos en ella [4]. No obstante, estas últimas extraen información y la resumen, pero no generan un discurso desde cero. Relacionados directamente con la generación de descripciones de empresas o textos para realizar auditorías no se ha encontrado ningún estudio.

Por estos motivos, se ha desarrollado una herramienta, TextReportGenerator, que permite la generación de descripciones de empresas ficticias con base a una serie de datos iniciales insertados previamente, con el objetivo de realizar una práctica de auditoría de protección de datos personales sobre las mismas. Se otorga variabilidad a las descripciones para evitar plagio entre los alumnos. Por otro lado, las descripciones mejoran la completitud y homogeneidad sobre las creadas manualmente, ahorrando tiempo al profesorado y disminuyendo la necesidad de consultar dudas de los alumnos acerca del enunciado de la práctica.

## 2. Solución propuesta

La herramienta desarrollada se muestra en la Figura 1 (izquierda), siendo esta la pestaña principal (Generador), dónde se encuentra el botón para cargar la base de datos (BDD). De aquí se extraerán los datos para las descripciones y las diferentes opciones.

botón «Generar texto» se activa al cargar la base de datos, generando una descripción de empresa diferente cada vez que se activa el botón. Se pueden incluir las cuentas de correo a las que se enviará la descripción de empresa generada (pulsando Enviar, con opción CCO y CC), indicando la contraseña del servidor de correo (que no se guarda).

Para crear la BDD con la información necesaria para generar las descripciones, está la segunda pestaña (Caso de uso), mostrada en la Figura 1 (derecha). Aquí, se crea la BDD y se pueden incluir tantas entradas (secciones) y opciones para cada sección como se desee añadiendo las filas necesarias. La introducción y las conclusiones son opcionales y sirven para que cuando aparezca esa sección en el texto generado lleve siempre ese inicio y/o final. Las opciones son alternativas de entre las que escogerá

una aleatoriamente, y que se colocará entre la introducción y/o conclusión de cada sección. A cada una de estas opciones se le puede dar un peso para que tenga mayor o menor probabilidad de salir. Si tienen el mismo peso, todas tendrán las mismas posibilidades de salir, pues el número indica el número de veces que se inserta esa opción en la «bolsa de aleatorios» (cada opción tiene una probabilidad de salir igual a su peso dividido por la suma de todos los pesos). También se ofrece la alternativa de «depende de», que permite que una opción dependa de alguna otra opción anterior, de tal manera que la opción solo podría ser elegida si previamente la opción de la que depende ya fue elegida. «dependencia» permite marcar una opción para indicar que otras van a depender de ella. Esto permite generar textos más ricos y con mayor variabilidad. No obstante, si falta o están mal configurados los «depende de» y/o la «dependencia», esto puede dar lugar a alguna incongruencia, ambigüedad o error. En este caso, se debe ir a la pestaña «caso de uso», para revisar y crear bien estas dependencias para futuras generaciones. Si el alumno detectara algún fallo, debe consultar al profesor para solventarlo. La forma de comprobar que todo se haya creado correctamente es generando informes y leyéndolos, y revisando las dependencias, pues las posibilidades pueden ser muchas, según el tamaño de la BDD creada.

En la pestaña «Configuración» se configura y guardan el correo del profesor, la configuración necesaria para el envío de correos (correo y servidor SMTP), título, encabezado y firma del correo, y las opciones de guardar, cancelar y prueba.

La pestaña «Acerca de...» tiene la información sobre los desarrolladores y el proyecto. La aplicación está internacionalizada al inglés.

TextReportGenerator puede ser utilizado libremente, pues cuenta con licencia GNU General Public

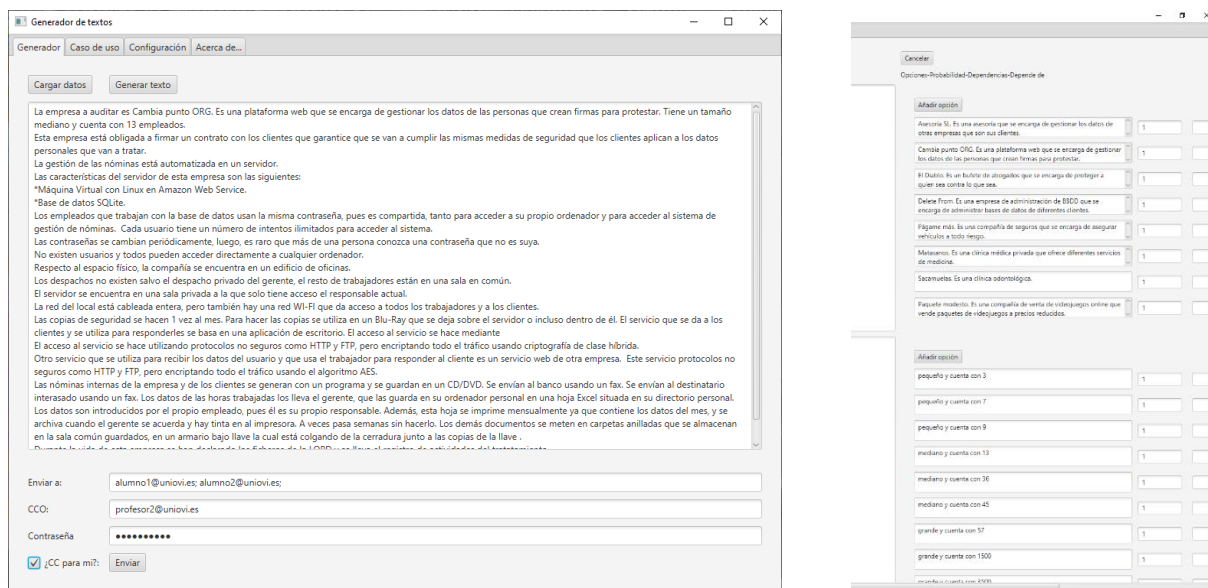


Figura 1 Pestaña principal de TextReportGenerator y parte de las opciones de la pestaña de creación de la BDD.

License v3.0. El código fuente del programa está en <https://github.com/gonzalezgarcia/TextReportGenerator>. En la carpeta «examples» está la base de datos de empresas ficticias utilizada «ASLEPI.json», el fichero de configuración propio (de prueba) «configuration.conf», y un PDF con diferentes ejemplos. La carpeta «release» contiene la última versión de la aplicación. La aplicación solo funciona en Windows de momento, y necesita la versión 8 de Java al menos.

### 3. Conclusiones

Se presenta la herramienta TextReportGenerator, que permite generar textos con descripciones de empresas ficticias para desarrollar ejercicios de auditoría de protección de datos personales, mediante la generación de enunciados de tipo descriptivo, con variabilidad, y de forma automática. Esta herramienta es novedosa pues, no se encuentran herramientas de este estilo en la literatura. Existen numerosas herramientas para educación que generan problemas matemáticos y algorítmicos por medio de variables parametrizables, o que crean huecos en textos, pero no de generación variable de textos completos descriptivos para enunciados de ejercicios con base en unas reglas.

La utilización de esta herramienta constata el ahorro del tiempo tanto para el profesor como para el alumno. Se requiere crear la base de datos, que lleva tiempo, pero tras esto, la generación del texto es inmediata. Luego, se ha reducido enormemente la dedicación del profesor a la generación de los enunciados. Además, se ha reducido el tiempo dedicado por alumnos y profesores a la resolución de dudas y aclaraciones por correo electrónico (se necesitaron hasta 5 veces menos correos que con los enunciados generados manualmente en años anteriores). La variabilidad introducida en los enunciados creemos que necesariamente ha reducido la posible tasa de plagio en las auditorías entregadas, además de que (con una base de datos de descripciones adecuada), la dificultad de los enunciados es más homogénea, lo cual es más justo para los alumnos.

Aunque la herramienta está destinada a la asignatura de ASLEPI, podría ser utilizada en asignaturas que traten contenidos similares, y en general, por cualquier asignatura en la que fuera necesario generar textos de enunciados descriptivos a partir de una serie de supuestos con diferentes alternativas.

### Agradecimientos

Este proyecto de innovación docente ha sido desarrollado en dos convocatorias de proyectos de innovación docente de la Universidad de Oviedo, cuyos códigos son PINN-18-A-96 y PINN-19-B-039, y en parte financiado por el Departamento de Informática de la Universidad de Oviedo.

### Referencias

- [1] Jefatura del Estado, “Ley Orgánica de Protección de Datos,” 2018. <https://www.boe.es/buscar/pdf/2018/BOE-A-2018-16673-consolidado.pdf> (acceso 10 Feb., 2022).
- [2] Chris Alvin, S. Gulwani, R. Majumdar, y S. Mukhopadhyay. Automatic Synthesis of Geometry Problems for an Intelligent Tutoring System. Oct. 2015. Disponible en <https://arxiv.org/abs/1510.08525>.
- [3] Omar Gil, R. Motz, y E. Sanchez. Exercise specification for collaboration, reusing and automatically generating evaluations. En *2017 Twelfth Latin American Conference on Learning Technologies (LACLO)*, pp. 1–4, Oct. 2017, doi: 10.1109/LACLO.2017.8120885.
- [4] Takuya Goto, T. Kojiri, T. Watanabe, T. Iwata, y T. Yamada. Automatic Generation System of Multiple-Choice Cloze Questions and its Evaluation. *Knowledge Management & E-Learning*, vol. 2, no. 3, pp. 210–224, 2010.
- [5] I-Han Hsiao, S. Sosnovsky, y P. Brusilovsky. Guiding students to the right questions: adaptive navigation support in an E-Learning system for Java programming. *Journal of Computer Assisted Learning*, vol. 26, no. 4, pp. 270–283, Jul. 2010, doi: 10.1111/j.1365-2729.2010.00365.x.
- [6] Thomas James Tiam-Lee y K. Sumi. Procedural Generation of Programming Exercises with Guides Based on the Student’s Emotion. En *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 1465–1470, Oct. 2018. doi: 10.1109/SMC.2018.00255.
- [7] Filiz Kurt-Karaoglu, N. Schwinning, M. Striwe, B. Zurmaar, y M. Goedicke. A Framework for Generic Exercises with Mathematical Content. En *2015 International Conference on Learning and Teaching in Computing and Engineering*, pp. 70–75, Apr. 2015, doi: 10.1109/LaTiCE.2015.11.
- [8] Ferrán Prados, I. Boada, J. Soler, y J. Poch. Automatic Generation and Correction of Technical Exercises. En *International Conference on Engineering and Computer Education (ICECE05)*, 2005. Disponible en <http://acme.udg.edu/articles/ICECE2005.pdf>
- [9] Akiyoshi Wakatani y T. Maeda. Automatic generation of programming exercises for learning programming language. En *2015 IEEE/ACIS 14th International Conference on Computer and Information Science (ICIS)*, pp. 461–465, Jun. 2015, doi: 10.1109/ICIS.2015.7166637.