

Trebial: un juego de preguntas generadas a partir de la DBpedia

Guillermo Vega-Gorgojo^{1,2}

¹ Don Naípe

² Departamento de Informática, Universidad de Oslo

Resumen Esta demostración presenta el Trebial, un juego de preguntas multijugador para dispositivos Android con más de 200K preguntas de opción múltiple (en español y en inglés) generadas a partir de la DBpedia. Las preguntas se han obtenido utilizando un proceso de extracción de datos y un generador de preguntas basado en plantillas. El sistema emplea un servidor para gestionar la colección de preguntas y las imágenes asociadas, mientras que se ha desarrollado una aplicación móvil y publicada en Google Play en marzo de 2017 con los nombres de ‘Trebial’ en español y ‘Clover Quiz’ en inglés. El juego se ofrece gratis y ha sido descargado por más de 5K usuarios hasta el momento en que se escribió este texto. Durante este periodo se han presentado más de 609K preguntas y la valoración del juego es de 4,3 sobre 5,0. De esta manera, el Trebial demuestra las ventajas de las tecnologías semánticas para extraer datos y automatizar la generación de preguntas de opción múltiple de manera escalable.

1 Introducción

DBpedia es una enorme base conocimiento multilingüe que constituye el esqueleto de la Web Semántica. Una de las aplicaciones más atractivas del uso de la DBpedia es la generación de preguntas para juegos tipo trivial. En la literatura pueden encontrarse algunos intentos preliminares [1,2,3], pero estas iniciativas no han tenido mucho recorrido debido a las limitaciones de los mecanismos de generación de preguntas empleados que no han sido capaces de producir preguntas variadas, numerosas y entretenidas. Más en concreto, los tipos de preguntas permitidos son muy limitados, los conjuntos de preguntas generados son bastante pequeños (del orden de miles) y no parece que tengan usuarios. Además, algunos de estos trabajos obtienen las preguntas formulando consultas en vivo a la DBpedia con lo que la latencia es demasiado grande para servir en un juego interactivo, tal como se indica en [2].

La hipótesis de este trabajo es que la generación de preguntas a partir de la DBpedia puede mejorarse de manera significativa dividiendo este proceso en una fase de extracción de datos y una de generación de preguntas que sea versátil. Siguiendo esta aproximación y especificando de manera declarativa clases y plantillas de preguntas de un dominio de interés pueden producirse preguntas variadas, numerosas y de alta calidad. Las preguntas así conseguidas pueden

alojarse en un servidor que cumpla con los requisitos de latencia de un juego interactivo de preguntas. El caso de aplicación es un juego por turnos en el que dos jugadores compiten en un tablero con forma de trébol respondiendo a preguntas de diferentes dominios. Esta demostración presenta dicho juego, llamado Trebial, y que está disponible para dispositivos Android a través de Google Play.

2 Extracción de datos y generación de preguntas

Puesto que el juego está orientado a dispositivos móviles, el Trebial utiliza preguntas de opción múltiple con 4 opciones para reducir la escritura en la app. Para generar las preguntas, los datos de interés se extraen de DBpedia. En una primera etapa se produce un fichero de especificación con las clases del dominio de interés, por ejemplo *Museo* o *Cuadro* in Arte. Cada clase se asocia con una consulta SPARQL para recuperar las entidades correspondientes. El fichero de especificación identifica también los literales que deben extraerse para las entidades de las clases objetivo –como etiquetas, años o URLs de las imágenes– así como las propiedades entre entidades de distintas clases. A continuación, un *script* utiliza un fichero de especificación como entrada y consulta DBpedia de manera sistemática para recuperar los datos disponibles en el dominio correspondiente. En concreto, el *script* extrae las entidades pertenecientes a cada clase, sus literales y las relaciones con otras entidades. La salida de esta fase es un fichero con un objeto JSON para cada entidad encontrada.

El proceso de extracción de datos incluye una fase de anotación de categorías basada en las categorías de Wikipedia que se han extraído para cada entidad. Los “wikipedistas” anotan los artículos de la Wikipedia con categorías organizadas de manera jerárquica que refleja la noción de “ser una subcategoría de”. Puesto que hay más de un millón de categorías en la Wikipedia, un *script* extrae de manera automática un subconjunto relevante para el dominio de interés. Esto se hace formulando consultas SPARQL para extraer las categorías con las que están anotadas las entidades obtenidas en la fase anterior y posteriormente obteniendo el conjunto extendido de categorías. Finalmente, otro *script* toma como entrada los datos extraídos, las categorías extendidas, y las categorías de interés para las preguntas, por ejemplo `dbc:Baroque_paintings` (cuadros barrocos) para la clase *Cuadro*. Para cada categoría candidata, el *script* obtiene la lista expandida de subcategorías y anota las entidades apropiadas de la clase objetivo.

Con los ficheros de anotaciones obtenidos se generan las preguntas utilizando plantillas multilingües. Las plantillas actúan sobre las entidades de una sola clase o de dos, dependiendo del tipo. Los tipos de plantillas disponibles son los siguientes: **imagen**, p. ej. *¿Cuál es el cuadro de la imagen?*; **booleana**, p. ej. *¿Cuál es el edificio modernista?*; **booleana negativa**, p. ej. *¿Quién NO es un escultor griego clásico?*; **grupo**, p. ej. *¿Cuál es el estilo artístico del pintor {{pintor.nombre}}?* (opciones: *Gótico, Renacimiento, Barroco, Manerismo, Romanticismo*); **fecha**, p. ej. *¿Cuándo nació el pintor {{pintor.nombre}}?*; **mayor**, p. ej. *¿Qué país tiene mayor población?*; **numérico**, p. ej. *¿Cuál es la población de*

$\{\{ciudad.nombre\}\}?$; **relación**, p. ej. *¿Quién es el pintor de “ $\{\{cuadro.nombre\}\}?$ ”?*; y **relación negativa**, p. ej. *¿Qué castillo NO está en $\{\{país.nombre\}\}?$?*

Un *script* se encarga de generar las preguntas a partir de las plantillas. De manera resumida, el *script* primero obtiene el conjunto de entidades que cumplen con las restricciones de la plantilla, p. ej. cuadros con animales del periodo barroco. Entonces generará una pregunta para cada entidad extrayendo la URL de la imagen (para apoyar la pregunta) y el nombre del cuadro (ésta será la respuesta correcta). Finalmente, se prepara la lista de opciones erróneas para cada pregunta generada. La Figura 1(arriba) muestra dos preguntas de ejemplos obtenidas con este proceso. Además, se emplea un estimador de dificultad para ajustar la dificultad de las preguntas a las habilidades de los jugadores. En la Tabla 1 se incluyen las cifras agregadas de las preguntas generadas para los diferentes dominios.

Tabla 1: Resumen del proceso de generación de preguntas para los dominios del Trebial.

	Animales	Arte	Libros	Cine	Geo	Música	Tecnología	TOTAL
# de plantillas	125	269	387	295	724	767	374	2,941
# de preguntas (español)	15,342	18,121	23,580	49,208	24,086	36,136	21,014	187,487
# de preguntas (inglés)	15,347	27,523	46,403	50,199	24,484	36,075	21,017	221,048

3 Demostración

El conjunto de preguntas generado se ha insertado en una base de datos MongoDB. Node.js se ha empleado para tratar las peticiones de preguntas, mientras que se ha configurado un servidor Nginx para alojar las imágenes del juego (37K imágenes de baja resolución, 1.12 GB en total) y para redirigir las peticiones de preguntas a Node.js. Con respecto a la aplicación móvil, he desarrollado una versión Android del juego que puede utilizarse en móviles y tabletas –ver las capturas de pantalla de la Figura 1. La app móvil permite partidas de dos jugadores, modo de juego individual, estadísticas, marcadores e hitos.

El juego para Android se lanzó el 11 de marzo de 2017 con los nombres de ‘Trebial’ en español y ‘Clover Quiz’ en inglés. El juego se ofrece gratis en Google Play en url<https://play.google.com/store/apps/details?id=donnaipe.trebial> y es parte del catálogo de Don Naípe, una compañía especializada en juegos de cartas para dispositivos móviles. El Trebial se ha promocionado con una campaña interna de publicidad lanzada del 13 al 16 de marzo, es decir, se mostraron anuncios del Trebial en otros juegos Android de Don Naípe¹. En el momento en que

¹ <https://play.google.com/store/apps/developer?id=Don+Naípe>

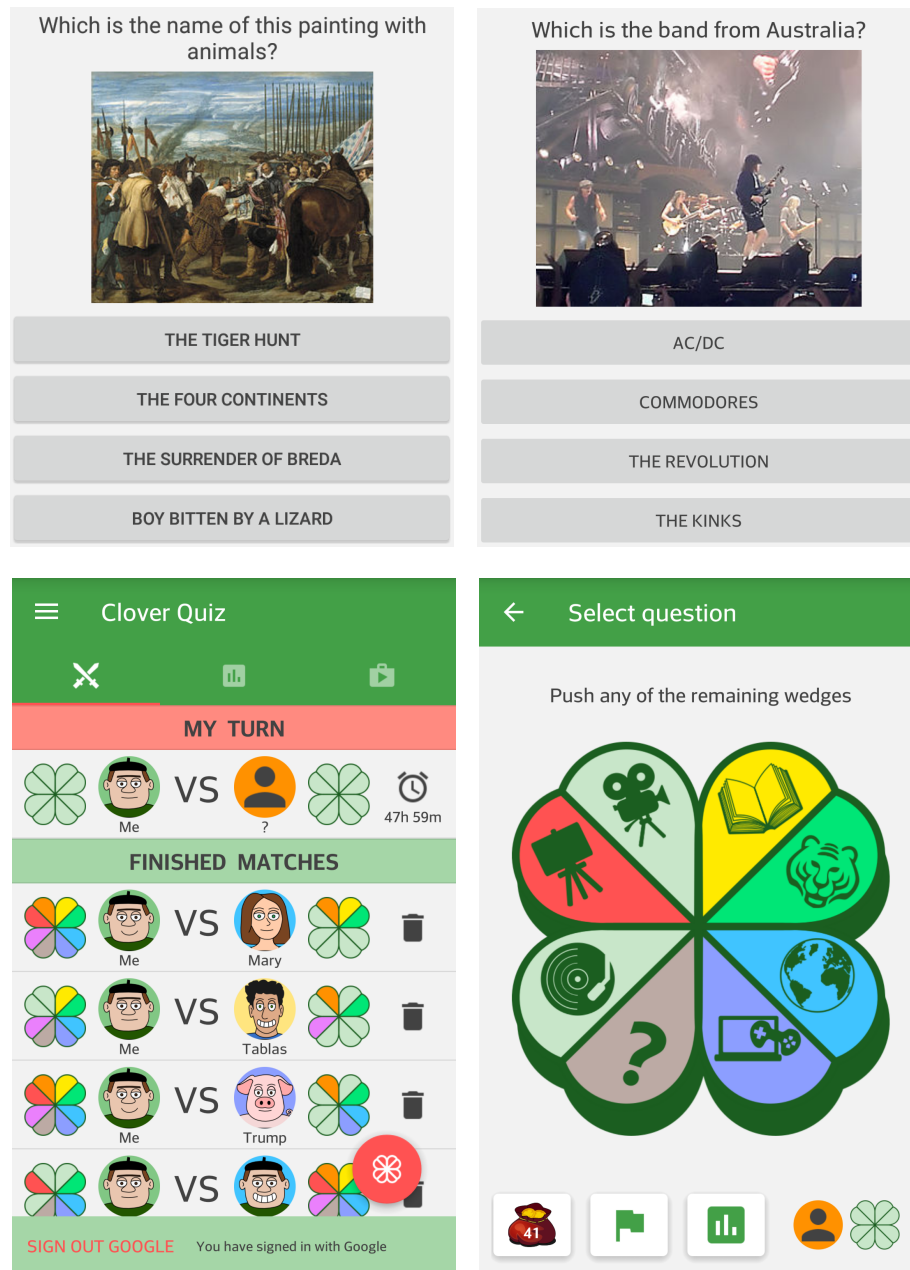


Figura 1: Preguntas de ejemplo para los dominios de las Artes y Animales (arriba). Capturas de pantalla del juego de Trebial (abajo).

se escribió este texto (julio de 2017) más de 5K usuarios se han descargado el juego y contestado a más de 609K preguntas. Los usuarios del Trebial también han dado realimentación a través de Google Play; en concreto, la valoración media es de 4,3 sobre 5,0 y los comentarios de los usuarios son generalmente muy alentadores.

Agradecimientos

Este trabajo ha sido parcialmente financiado por el Norwegian Research Council a través del centro de innovación SIRIUS (NFR 237898).

Referencias

1. Bratsas, C., Chrysou, D.E., Eftychiadou, E., Kontokostas, D., Bamidis, P., Antoniou, I.: Semantic Web game based learning: An i18n approach with Greek DBpedia. In: Proceedings of the 2nd LiLe Workshop. Lyon, France (Apr 2012)
2. Mütsch, F.: Auto-generated trivia questions based on DBpedia data (Feb 2017), URL: <https://github.com/n1try/linkedata-trivia>, last accessed July 2017
3. Waitelonis, J., Ludwig, N., Knuth, M., Sack, H.: Whoknows? Evaluating linked data heuristics with a quiz that cleans up DBpedia. *Interactive Technology and Smart Education* 8(4), 236–248 (2011)