Clover Quiz: a Mobile Trivia Game Based on DBpedia Data

Guillermo Vega-Gorgojo[1,2]

[1] Don Naipe
[2] Department of Informatics, University of Oslo

Abstract. This demonstration presents Clover Quiz, a turn-based multiplayer trivia game for Android devices with more than 200K multiple choice questions (in English and Spanish) about different domains generated out of DBpedia. Instead of live queries, questions are created offline through a data extraction pipeline and a versatile template-based mechanism. A back-end server manages the question set and the associated images, while a mobile app has been developed and released in Google Play on March 2017 under the names 'Clover Quiz' in English and 'Trebial' in Spanish. The game is available free of charge and has been downloaded by more than 5K users at the time of this writing. During this period, more than 609K questions have been posed, and the overall rating of the game is 4.3 out of 5.0. Clover Quiz thus demonstrates the advantages of semantic technologies for collecting data and automating the generation of multiple choice questions in a scalable way.

1 Introduction

DBpedia is a large-scale and multilingual knowledge base that constitutes the main hub of the Semantic Web. An appealing application case is the generation of questions for trivia games from DBpedia. Some preliminary attempts can be found in the literature [1,2,3], but these initiatives have fallen short due to simple question generation schemes that are not able to produce varied, large, and entertaining questions. Specifically, supported question types are rather limited, reported sizes of the generated question sets are relative low (in the range of thousands), and no user base seems to exist. Moreover, some of these works create the questions by submitting live queries to the public DBpedia endpoint, hence latency is too high for an interactive trivia game, as reported in [2].

The hypothesis is that creating questions from DBpedia can be significantly improved by splitting this process in a data extraction and a versatile question generation stages. This approach can produce varied, numerous, and high-quality questions by declaratively specifying the classes and question templates of the domains of interest. The generated questions can then be hosted in a back-end server that meets the latency requirements of an interactive trivia game. The target case is a turn-based multiplayer trivia game in which two players compete over a clover-shaped board by answering multiple choice questions from different domains. This demonstration presents such a game, named Clover Quiz, that is available for Android devices through Google Play.

2   Data Extraction and Question Generation

Since the game is purposed for mobile devices, Clover Quiz employs multiple choice questions with 4 options to limit user typing. To generate the questions, the data of interest is gathered from DBpedia. In a first stage, a domain specification file is authored with the classes in the domain of interest, e.g. `Museum`, or `Painting` in Arts. Each class is associated with a SPARQL query for retrieving the corresponding entities. The specification file also identifies the literals to be extracted for the entities of a target class –like labels, years, or image URLs– as well as relations between entities of different classes. Next, a script takes a specification file as input and systematically queries DBpedia to retrieve the data available of the target domain. Specifically, this script gathers the entities belonging to each class, their literals, and their relations with other entities. The output of this stage is a file with a JSON object for every entity found.

The data extraction pipeline includes a category annotation stage based on the Wikipedia categories extracted for each entity. Wikipedia contributors annotate articles with suitable categories that are organised into a hierarchy that reflects the notion of "being a subcategory of". Since there are more than one million Wikipedia categories, a script automatically gathers a relevant subset of the domain of interest by posing SPARQL queries to retrieve the annotated categories of the extracted entities and then the expanded category set. Finally, another script takes as input the extracted data, the expanded categories, and the Wikipedia categories of interest, e.g. `dbc:Baroque_paintings` for the `Painting` class. For every candidate category, the script obtains the expanded list of subcategories and annotates the appropriate entities of the target class.

With the obtained annotated data files, questions are then generated using multilingual question templates. A template can involve either entities of a single class or of two classes, depending on the template type. These are the following: `image`, e.g. *Which is the painting of the image?*; `boolean`, e.g. *Which is the modernist building?*; `boolean negative`, e.g. *Which is NOT an Ancient Greek sculptor?*; `group`, e.g. *Which is the artistic style of the painter {{painter.label}}?* (options: *Gothic, Renaissance, Baroque, Mannerist, Romantic*); `date`, e.g. *When was {{painter.label}} born?*; `greatest`, e.g. *Which country has the largest population?*; `numeric`, e.g. *Which is the population of {{city.label}}?*; `relation`, e.g. *Who is the painter of "{{painting.label}}"?*; and `relation negative`, e.g. *Which castle is NOT in {{country.label}}?*.

A script is in charge of generating the questions by evaluating every template. Briefly, the script first obtains the set of entities that comply with the template requirements, e.g. paintings with animals from the Baroque period. It will then generate a question for each occurrence by getting the image URL (to support the question) and the label of the painting (this will be the correct answer). Finally, a list of distractors is prepared from the target set. Fig. 1(top) shows two sample questions generated through this process. Although not described here due to space constraints, a question difficulty estimator is employed in order to adjust the difficulty of the questions to the player skills. Table 1 presents some aggregated figures of the question set obtained for the target domains.

Table 1: Summary of the question generation process for the different domains.

	Animals	Arts	Books	Cinema	Geo	Music	Tech	TOTAL
# of templates	125	269	387	295	724	767	374	2,941
# of questions (Spanish)	15,342	18,121	23,580	49,208	24,086	36,136	21,014	187,487
# of questions (English)	15,347	27,523	46,403	50,199	24,484	36,075	21,017	221,048

3 Demonstration Overview

The obtained question set was deployed in a MongoDB database. Node.js was employed to handle question requests, while an Nginx server was configured to host the game images (37K low-resolution images, totalling 1.12 GB) and to forward question requests to Node.js. As for the mobile app, I coded an Android version of the game that can be played in phones and tablets –see sample snapshots in Fig. 1. The mobile app supports two-player matches, single-player mode, statistics, leaderboards and achievements.

The game was released for Android devices on March 11, 2017 under the names 'Clover Quiz' in English and 'Trebial' in Spanish. It is available free of charge through Google Play at `https://play.google.com/store/apps/details?id=donnaipe.trebial` and is part of the catalogue of Don Naipe, a sole proprietorship company specialized in Spanish card games for mobile devices. Clover Quiz was promoted with an in-house ad campaign that ran from March 13 to March 16, i.e. other Android games by Don Naipe[1] showed interstitial ads about Clover Quiz. At the time of this writing (July 2017), more than 5K users have downloaded the game, requesting more than 609K questions. Clover Quiz users have also given feedback through Google Play. Specifically, the average rating is 4.3 out of 5.0 and users' comments are generally very supportive.

Acknowledgements

References

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/linkeddata-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)

---

[1] `https://play.google.com/store/apps/developer?id=Don+Naipe`

Which is the name of this painting with animals?

THE TIGER HUNT

THE FOUR CONTINENTS

THE SURRENDER OF BREDA

BOY BITTEN BY A LIZARD

Which is the band from Australia?

AC/DC

COMMODORES

THE REVOLUTION

THE KINKS

Clover Quiz

MY TURN

Me VS ? 47h 59m

FINISHED MATCHES

Me VS Mary

Me VS Tablas

Me VS Trump

Me VS

SIGN OUT GOOGLE   You have signed in with Google

Select question
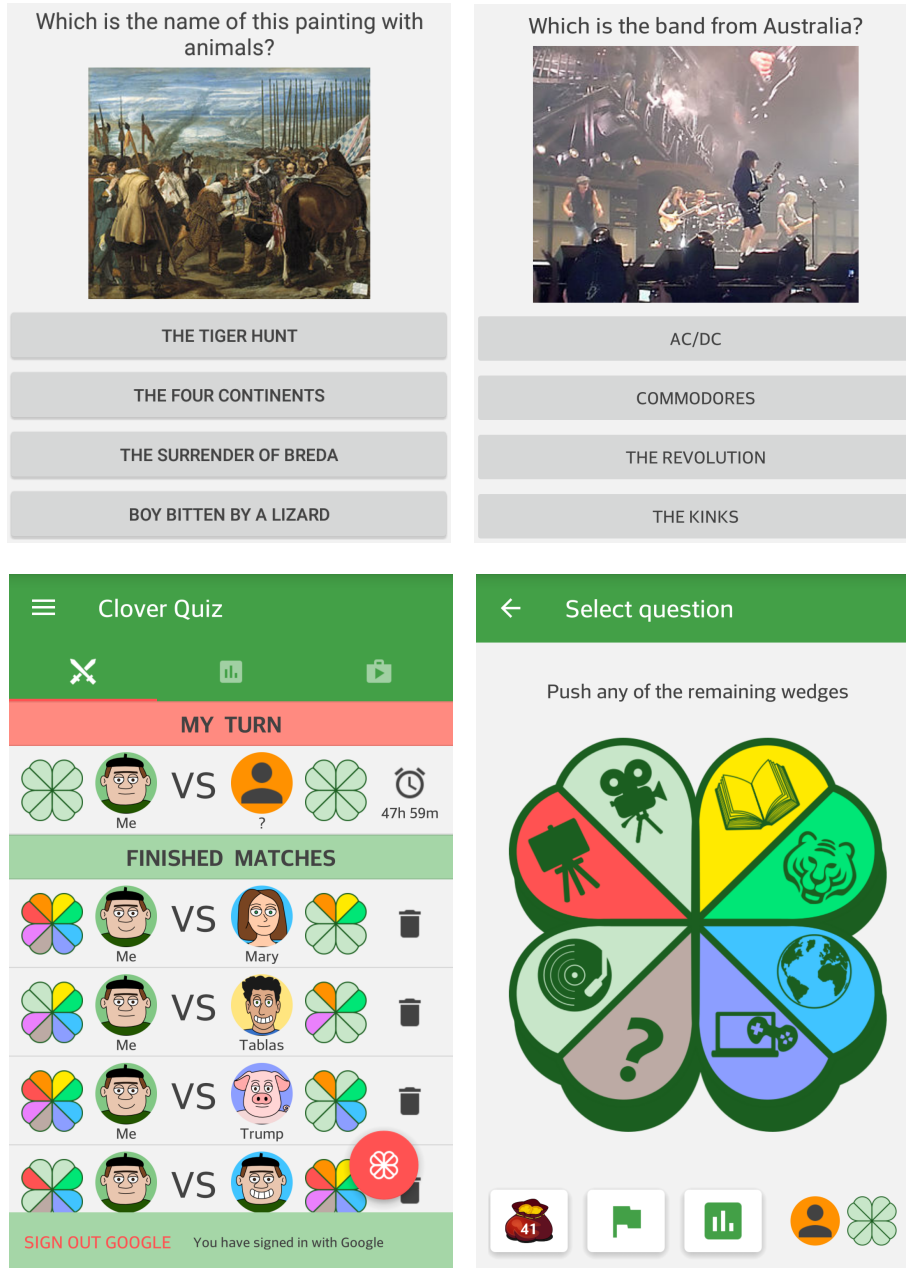
Push any of the remaining wedges

41

Fig. 1: Sample questions from the Arts and Animals domains (top). Sample snapshots of the main and game screens of Clover Quiz (bottom).