

Tabouid: un jeu de langage et de culture générale généré à partir de Wikipédia

Timothée Bernard

Laboratoire de linguistique formelle, Université de Paris, France
timothee.bernard@u-paris.fr

RÉSUMÉ

Nous présentons des résumés en français et en anglais de l'article (Bernard, 2020), présenté lors de la conférence *58th Annual Meeting of the Association for Computational Linguistics* (ACL 2020). L'article détaille comment un éventail de techniques relativement simples de TAL et d'apprentissage automatique peuvent être combinées pour générer à partir de Wikipédia le contenu d'un jeu de langage et de culture générale. L'article peut être vu comme définissant un projet stimulant pour des étudiant-e-s en TAL et le jeu lui-même a effectivement été implémenté sous la forme de Tabouid, une application Android et iOS.

ABSTRACT

Tabouid: a Wikipedia-based word guessing game

We present abstracts in English and in French for (Bernard, 2020), a paper that was presented at the 58th Annual Meeting of the Association for Computational Linguistics (ACL 2020). The paper details how a range of relatively simple NLP and machine-learning techniques can be brought together to automatically generate from Wikipedia the content of a word-guessing game. The paper can be seen as defining an engaging project for students enrolled in NLP programs and the game itself has been implemented as Tabouid, an Android and iOS application.

MOTS-CLÉS : jeu, Wikipédia, génération automatique de contenu, application, projet pédagogique.

KEYWORDS: game, Wikipedia, automatically generated content, application, pedagogical project.

1 Résumé en français

Nous présentons Tabouid, un jeu de langage et de culture générale généré automatiquement à partir de Wikipédia. Le but du jeu consiste à faire deviner aux autres joueurs et en un minimum de temps le titre d'une *carte* sans prononcer les mots interdits qu'elle spécifie. Tabouid contient 10 000 cartes (virtuelles) en anglais et autant en français. Ces cartes traitent non seulement de mots ou d'expressions courantes, mais aussi de concepts issus d'un large spectre de domaines tels que les arts, l'histoire ou les sciences. Chacune des cartes correspond à un article Wikipédia et, inversement, tout article peut servir à générer une carte. Un éventail de techniques relativement simples de TAL (racinisation, sélection par comptage, filtrage par règles, etc.) et d'apprentissage automatique (*scoring*, *active learning*, etc.) sont combinées efficacement pour former un algorithme en deux temps. Premièrement, une grande quantité d'articles Wikipédia sont évalués avec un score allant de 0 à 1 — ce score estime la difficulté, ou, vu autrement, la jouabilité de l'article. Ensuite, les meilleurs articles sont convertis en carte : il s'agit d'extraire, pour chacun d'entre eux, une liste de mots ou expressions interdites.

Toutes les cartes de Tabouid étant associées à un score de jouabilité, la difficulté du jeu peut être réglée à l'aide d'un paramètre de seuil. Un tel réglage permet d'adapter le jeu à des publics divers, incluant par exemple enfants ou locuteurs non natifs, dont la maîtrise linguistique ou la connaissance de la culture associée à la langue sélectionnée peut varier. De plus, nous croyons que cet article à une certaine valeur pédagogique en définissant un projet d'implémentation pour des étudiant-e-s en TAL/linguistique computationnelle.

Le jeu est disponible sous la forme d'une application Android (<https://play.google.com/store/apps/details?id=com.tot.tabouid>) et iOS (<https://apps.apple.com/us/app/tabouid/id1477994156>) entièrement gratuite et ne contenant aucune publicité. Les cartes ont été pré-générées et incluses dans l'application ; aucune connexion internet n'est requise durant le jeu.

2 Abstract in English

We present Tabouid, a word-guessing game automatically generated from Wikipedia. During a turn of the game, one player tries to make the other players guess the title of *card* as quickly as they can and without using any of the banned words mentioned on the card. Tabouid contains 10,000 (virtual) cards in English, and as many in French, covering not only words and linguistic expressions but also a variety of topics including artists, historical events or scientific concepts. Each card corresponds to a Wikipedia article, and conversely, any article could be turned into a card. A range of relatively simple NLP (stemming, count-based selection, rule-based filtering, etc.) and machine-learning (scoring, active learning, etc.) techniques are effectively integrated into a two-stage process. First, a large subset of Wikipedia articles are scored — this score estimates the difficulty, or alternatively, the playability of the page. Then, the best articles are turned into cards by selecting, for each of them, a list of banned words based on its content.

All cards in Tabouid are associated with a difficulty score. This allows the difficulty level of the game to be set in a straightforward way. With such an adaptable difficulty, the game can accommodate various groups of players, which could include individuals such as children or foreigners, whose level of proficiency or knowledge of the culture associated with the target language may vary. In addition, we believe that this paper can have some pedagogical value by defining an implementation project for students enrolled in NLP or computational linguistic programs.

The game is available as an Android (<https://play.google.com/store/apps/details?id=com.tot.tabouid>) and iOS (<https://apps.apple.com/us/app/tabouid/id1477994156>) application that is entirely free and does not contain any advertisement. The cards have been pre-generated and packaged with the application; an internet connection is not required during the game.

Références

BERNARD T. (2020). Tabouid : a Wikipedia-based word guessing game. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics : System Demonstrations*, p. 24–29 : Association for Computational Linguistics.