

Analyse thématique comparative des discours politiques et de leur diffusion dans le Wikipédia francophone

Jeanne Vermeirsche¹, Eric SanJuan², Tania Jiménez³, Christèle Lagier⁴

¹JPEG Avignon Université – jeanne.vermeirsche@univ-avignon.fr

²LIA Avignon Université – eric.sanjuan@univ-avignon.fr

³LIA Avignon Université – tania.jimenez@univ-avignon.fr

⁴JPEG Avignon Université – christele.lagier@univ-avignon.fr

Abstract

This work involves analyzing the themes of press releases from political parties using the Latent Dirichlet Allocation (LDA) method and studying how these themes are represented in articles on French-language Wikipedia. The objective is to determine if Wikipedia maintains neutrality towards political positions. The project begins with the collection of press releases from various political parties and Wikipedia articles related to these parties or political themes. After a thorough preprocessing of the texts, LDA is applied to identify the major themes in the communiqués of each party. A comparative analysis is then conducted to assess whether the identified themes are present in Wikipedia and if their treatment is neutral, without favoritism towards any particular party. This analysis combines quantitative and qualitative methods.

Keywords: Topic Analysis ; Stochastic Models ; LDA ; Political Discourse ; Wikipedia

Résumé

Ce travail consiste à analyser les thématiques des communiqués de presse des partis politiques en utilisant la méthode Latent Dirichlet Allocation (LDA). Nous étudions comment ces thématiques sont représentées dans les articles de Wikipédia francophone. L'objectif est de déterminer si Wikipédia maintient une neutralité vis-à-vis des positions politiques. Le projet débute par la collecte des communiqués de presse des partis politiques et des articles de Wikipédia relatifs à ces partis ou aux thématiques politiques. Après un pré-traitement minutieux des textes, le LDA est appliqué pour identifier les thèmes majeurs dans les communiqués de chaque parti. Une analyse comparative est ensuite menée pour évaluer si les thèmes identifiés sont présents dans Wikipédia et si leur traitement est neutre, sans favoritisme envers un parti en particulier. Cette analyse combine des méthodes quantitatives et qualitatives.

Mots clés : Analyse thématique ; Modèles Stochastiques ; LDA ; Discours politique ; Wikipédia