

Méthodologie de la Vitrine Démocratique

Centre d'analyse des Politiques Publiques (CAPP)

Accès et structuration des données

Chaque dix minutes, le contenu des unes médiatiques de 13 médias Québécois et Canadiens est récolté. Les quotidiens sont les suivants : CBC, CTV, Le Devoir, Global News, The Globe and Mail, Le Journal de Montréal, Montreal Gazette, National Post, La Presse, Radio-Canada, Toronto Star, TVA et Vancouver Sun. Le contenu des unes et leurs métadonnées sont entreposées dans une base de données structurée et sécurisée.

Mise à jour de la Vitrine démocratique

Les analyses sont disponibles pour trois périodes différentes. Les analyses quotidiennes sont ajustées trois fois par jour : le matin, après l'heure du dîner et le soir. Les analyses hebdomadaires et mensuelles sont ajustées chaque soir.

En tout temps, les analyses quotidiennes et hebdomadaires se basent sur les données des 7 derniers jours. Les analyses mensuelles celles des 30 derniers jours.

Reflet

Le contenu des unes médiatiques est segmenté en phrases, puis analysé par un dictionnaire d'enjeux composé des 12 grandes catégories de politiques publiques du CAPP. Chaque phrase est classifiée et reçoit un score quantifié selon l'enjeu abordé. Le contenu est ensuite divisé par enjeux, et la base de donnée filtrée pour conserver seulement le texte associé à au moins un enjeu.

Chaque phrase est ensuite soumise à l'API d'OpenAI. Une requête est formulée au modèle GPT-4o, pour formuler un résumé en trois points des phrases du jour et de la semaine, pour chacun des 12 enjeux. Le résumé du mois est produit similairement, en fournissant les résumés de la semaine, afin de produire un résumé mensuel.

Graphique d'enjeux

Le contenu des unes médiatiques est segmenté en phrases, puis analysé par un dictionnaire d'enjeux composé des 12 grandes catégories de politiques publiques du CAPP. Chaque phrase est classifiée et reçoit un score quantifié selon l'enjeu abordé. Le score consiste en une proportion des mentions d'un enjeu donné lors d'une période de temps par rapport aux mentions de l'ensemble des enjeux sur la même période de temps.

Les scores sont ensuite agrégés par jour, permettant d'identifier les thèmes dominants dans le discours médiatique. En fonction de la période d'analyse (jour, semaine, mois), un graphique met en évidence les trois enjeux les plus discutés sur la période sélectionnée. Le graphique quotidien couvre une semaine, en affichant les trois enjeux principaux de la dernière journée. Le graphique hebdomadaire et le graphique mensuel couvrent respectivement les sept et les trente derniers jours, soulignant les trois enjeux dominants de chaque période. Enfin, une option de filtrage permet de mettre en évidence individuellement chacun des 12 enjeux.

Graphique des partis politiques

Le contenu des unes médiatiques est segmenté en phrases pour isoler chaque mention de parti dans son contexte. Les phrases sont ensuite classées par langue. Le corpus est analysé par un dictionnaire de partis politiques, permettant d'identifier les mentions de chaque parti en chambre, ainsi que par un dictionnaire de ton, attribuant à chaque phrase un score de tonalité (positif, neutre ou négatif).

Les mentions des partis politiques sont pondérées pour donner plus d'importance aux événements récents. Dans le graphique quotidien, les mentions des dernières 24 heures reçoivent un poids maximal, suivi de pondérations décroissantes pour les six jours précédents, à partir de 25 %. Le graphique hebdomadaire applique également un poids maximal aux 24 dernières heures, puis réduit progressivement la pondération pour chacun des six autres jours de la semaine. Enfin, le graphique mensuel diminue la pondération sur une période de 30 jours.

Ce processus offre une représentation dynamique de la visibilité médiatique et de la tonalité associée à chaque parti.