

Informations sur la méthodologie

Introduction

Le [Radar de Parité](#) est un outil unique d'analyse de données qui mesure le ratio d'hommes et de femmes cités.es dans les médias d'information canadiens influents. L'agrégation de ce jeu de données offre des preuves tangibles qui enrichiront considérablement notre compréhension de l'ampleur et de l'impact de la représentation des femmes dans le discours public. Ce projet est le fruit d'une collaboration entre [Informed Opinions](#) et [Simon Fraser University](#). La recherche et le développement sont menés par le [Discourse Processing Lab](#).

L'équipe

Notre équipe technique à l'Université Simon Fraser est composée de chercheur.euses, de développeur.euses de logiciels et d'expert.es en sciences des données.

- Maite Taboada, chercheuse principale et directrice du Discourse Processing Lab
- Fatemeh Torabi Asr, chercheuse postdoctorale et gestionnaire de projet
- Mohammad Mazraeh, développeur de logiciels et ingénieur en apprentissage automatique
- Vagrant Gautam, linguiste informaticien
- Alexandre Lopes, spécialiste des données et gestionnaire de bases de données
- Junette Dianne Gonzales, annotatrice de données linguistiques
- Prashanth Rao, ingénieur principal en IA et en données
- Valentin-Gabriel Soumah, spécialiste français du NLP
- Philipp Eibl, développeur de logiciels et spécialiste du NLP
- Jillian Anderson, spécialiste du Big Data
- Philip Chen, administrateur de la base de données

Informed Opinions est soutenue par les développeurs web de Kanopi pour fournir l'interface visuelle numériques.

Processus d'analyse textuelle

Nous extrayons le texte et les métadonnées de tous les articles de presse publiés chaque jour sur le site Web de six grands médias canadiens de langue française. Nous traitons ensuite le texte en utilisant une variété de techniques de traitement du langage naturel et d'apprentissage automatique afin d'identifier les personnes citées et leur genre.

Ce processus comprend l'analyse syntaxique, la reconnaissance des entités nommées, l'extraction des citations et l'identification du genre, le tout en temps réel. L'ensemble des données et des annotations sont conservées sur les serveurs de la base de données de SFU. L'interface montre la proportion d'hommes et de femmes cité.es quotidiennement.

Orientations futures

Le Radar de Parité est un travail en cours. Depuis la sortie de la version originale anglophone, le [Gender Gap Tracker](#), en février 2019 et de la version francophone en mars 2022, nous avons continué à raffiner ses capacités et son degré de précision. Identifier les personnes citées dans les articles, agréger ces mentions pour identifier adéquatement plusieurs citations provenant d'un.e même intervenant.e et identifier correctement leur genre sont des tâches analytiques complexes. Il arrive à l'occasion que les outils de traitement du langage naturel n'identifient pas une citation ou mégenrent une personne citée.

Pour identifier le genre des intervenant.es, nous employons des bases de données numériques qui associent les noms à un genre. Ces bases de données ont leurs propres limites. Néanmoins, nos évaluations montrent que, dans un échantillon suffisamment large de textes annotés manuellement, la proportion d'hommes et de femmes cité.es calculée par le Radar de Parité s'approche des chiffres réels, avec une marge d'erreur de 5%. Nous travaillons constamment à réduire cette marge d'erreur, mais il importe aussi de noter qu'aucun média ne risque d'être injustement affecté.

Pour plus d'informations, vous pouvez consulter [Three years of monitoring gender representation in the media](#) (en anglais seulement).

Pour plus d'informations sur la méthodologie de la recherche, prière de contacter :

Maite Taboada mtaboada@sfu.ca

Directrice, Laboratoire de traitement du discours

Université Simon Fraser