

MARATHON DU WEB ÉDITION 2024



Master – Parcours communication numériques et organisations

FEMME & MÉDIAS

Projet EchoFem

Marathon du Web Édition 2024

Université Paul Valéry Montpellier 3
Faculté d'Éducation Sciences LLASHS UFR 6
Master MIASHS

Réalisé par :

Alya ZOUZOU
Alvin VEDEL
Jamila Cherkaoui
Dina El Hijjawi



March 6, 2024

Introduction

Ce document présente les avancements quotidiens du projet "Echo Fem", développé dans le cadre du marathon du web avec pour objectif de créer une interface web en quatre jours sur le thème "Femmes et Médias".

Lundi

2.1 Gestion du Projet sur GitHub et Trello

La gestion efficace de notre projet "Echo Fem" s'appuie sur l'utilisation combinée de GitHub et Trello, deux outils pour le contrôle de version, la collaboration et l'organisation des tâches.

2.1.1 GitHub pour le Contrôle de Version et la Collaboration

Le suivi de notre projet est assuré grâce à la plateforme GitHub, qui offre un système de contrôle de version et facilite la collaboration entre tous les membres de l'équipe. GitHub est notre outil principal pour suivre l'évolution du projet, intégrer des modifications, gérer les versions et améliorer la communication au sein de l'équipe. Pour accéder à notre répertoire GitHub et suivre les mises à jour en temps réel, veuillez visiter le lien suivant : https://github.com/AlvinVedel/femmes_et_medias.git.

2.1.2 Trello pour l'Organisation et la Gestion des Tâches

Parallèlement à GitHub, nous utilisons Trello pour l'organisation des tâches et la gestion du workflow du projet. Trello nous permet de créer des tableaux, des listes et des cartes pour organiser le travail et suivre l'avancement de chaque aspect du projet de manière visuelle. Pour consulter notre tableau Trello et voir l'organisation actuelle des tâches, veuillez visiter le lien suivant : <https://trello.com/b/YKuzTaE6>.

L'utilisation combinée de GitHub et Trello nous permet de maintenir une gestion de projet fluide en assurant une coordination entre les membres de l'équipe tout au long du développement de "Echo Fem".

2.2 Présentation du projet "Média et femmes"

Le projet "Echo Fem", développé dans le cadre du Marathon du Web Édition 2024, se positionne à l'intersection de la technologie, des études de genre et des médias. Ce projet vise à décortiquer et analyser la représentation des femmes dans les médias traditionnels, en s'appuyant sur une décennie de données issues des journaux télévisés de TF1, FR2, et FR3. Le choix de ces chaînes est stratégique, visant à embrasser un spectre large de la population et des perspectives médiatiques.

Objectifs du projet :

- **Analyse du discours médiatique** : À travers une approche multidimensionnelle, incluant l'analyse de sentiment, l'identification de mots-clés liés au féminisme et à l'intersectionnalité, ainsi que l'étude de la polarité des discours, nous visons à dévoiler les biais, stéréotypes et préjugés véhiculés.
- **Développement d'une interface web interactive** : Le cœur du projet "Echo Fem" réside dans la création d'une plateforme web qui permettra de visualiser de manière dynamique et interactive les résultats de nos analyses. Cette interface vise à rendre les données accessibles et compréhensibles par un large public, favorisant ainsi la prise de conscience et le débat autour des enjeux de représentation des femmes dans les médias.
- **Contribution au débat public** : Par la diffusion des résultats de notre enquête via un canal de diffusion important (réseaux sociaux et partenariat avec le Club de la Presse Occitanie).

Enjeux et impact sociétal :

Le projet "Echo Fem" s'inscrit dans une démarche de sensibilisation et d'éducation. En mettant en lumière les mécanismes de construction des images des femmes dans les médias, nous souhaitons contribuer à une prise de conscience collective sur l'importance de la diversité et de l'équité dans les représentations médiatiques. L'objectif ultime est de promouvoir un changement vers des pratiques médiatiques plus inclusives et respectueuses de la diversité des parcours et des identités féminines. En résumé, "Echo Fem" ne se veut pas seulement un projet d'analyse médiatique ; il aspire à encourager les médias à adopter une approche plus nuancée et équilibrée dans la représentation des femmes, essentielle pour la construction d'une société plus juste et égalitaire.

2.3 Présentation des données

Les données utilisées pour notre analyse sont issues des journaux télévisés de TF1, FR2, et FR3. Elles comprennent les méta-données suivantes pour chaque segment diffusé :

Colonne	Description
Titre	Le titre de l'article ou du segment télévisé.
Description	Résumé ou description détaillée du contenu.
Date	Date de publication ou de diffusion.
Order	Identifiant numérique pour l'ordre de présentation.
Presenter	Nom du présentateur ou animateur.
Editor	Éditeur ou responsable de l'article (si applicable).
URL	Adresse web de l'article ou segment complet.
URLTvNews	Adresse web de la diffusion télévisée associée.
ContainsWordGlobalWarming	Indique si le segment mentionne le réchauffement climatique.
Media	Nom du média de diffusion.
Month	Mois de publication ou diffusion.
Day	Jour de publication ou diffusion.

Table 1: Description des données utilisées pour l'analyse

2.4 Explication de notre démarche

Au cours de cette première journée de travail sur notre projet "Echo Fem", nous avons engagé des discussions enrichissantes avec plusieurs experts spécialisés dans l'analyse textuelle. Ces échanges nous ont permis d'identifier une méthodologie robuste pour notre analyse, inspirée en partie par les techniques décrites dans l'article trouvé sur le Gender Gap Tracker.

1. **Filtrage des données via LDA (Latent Dirichlet Allocation):** Nous commençons par appliquer l'algorithme LDA pour identifier les sujets prédominants dans nos données textuelles. Cette technique de modélisation de sujets nous aide à filtrer les données de manière à focaliser notre analyse sur les segments les plus pertinents concernant la représentation des femmes.
2. **Application de l'indice de parité de genre (Gender Parity Index):** À la suite du filtrage par LDA, nous calculons l'indice de parité de genre pour chaque sujet identifié. Cet indice nous permet d'évaluer la représentation des femmes de manière quantitative. Les résultats sont ensuite visualisés sous forme de cartes thermiques (heat maps), offrant une représentation claire de la parité de genre au sein des sujets analysés.
3. **Identification des mots-clés relatifs aux femmes via un clustering hiérarchique:** Nous employons des techniques de clustering hiérarchique pour regrouper les mots-clés associés aux femmes. Cette approche nous aide à comprendre comment les femmes sont décrites et dans quels contextes elles apparaissent.
4. **Analyse de sentiment:** Nous effectuons une analyse de sentiment sur les segments textuels liés aux femmes pour évaluer les tonalités positives, neutres, ou négatives associées à leur représentation.
5. **Analyse de la haine:** Parallèlement à l'analyse de sentiment, nous réalisons une étude spécifique sur les expressions de haine, identifiant les segments où les femmes sont ciblées par des discours haineux ou discriminatoires.
6. **Graphique évolutif du nombre d'articles concernant les femmes:** Nous construisons un graphique temporel pour tracer l'évolution du nombre d'articles mentionnant les femmes, permettant de visualiser les tendances au fil du temps.
7. **Création d'un dataset basé sur les prénoms féminins:** Nous constituons une base de données de prénoms féminins à travers la reconnaissance d'entités nommées et notre dataset afin de filtrer les articles mentionnant ou écrits par des femmes, enrichissant ainsi notre analyse.
8. **Liste des mots choquants liés aux femmes et visualisation graphique:** Nous compilons une liste de termes choquants ou dégradants utilisés pour décrire les femmes et les visualisons à travers différents graphiques pour mettre en évidence leur fréquence d'utilisation.
9. **Top 5 des mots les plus utilisés:** Nous identifions et analysons les cinq mots les plus fréquemment associés aux femmes dans notre corpus de données.
10. **Analyse de similitude:** Enfin, nous étudions les similitudes entre les différents segments textuels pour découvrir des patterns récurrents dans la représentation des femmes.

Cette démarche méthodologique nous permet d'aborder notre analyse avec une approche multidimensionnelle, combinant techniques quantitatives et qualitatives pour une compréhension approfondie de la représentation des femmes dans les médias. Le tout sera regroupé dans une page web effectuée à l'aide de Stream Lit.

Mardi

Pour approfondir notre analyse sur la représentation des genres dans les médias, nous avons mis en place une série d'analyses structurées.

1. Application de l'Analyse Latente de Dirichlet (LDA) et Calcul de l'Indice de Parité de Genre ajusté (GPI ajusté)

On commence par appliquer la LDA sur notre jeu de données complet. Cette technique nous aide à dégager les thèmes prédominants et les mots-clés récurrents dans les textes, révélant ainsi les groupes d'articles contenant des features similaires. Cela nous offre un aperçu des sujets abordés dans notre corpus.

Avec les thèmes et mots-clés identifiés, on calcule le GPI ajusté pour chaque thème, ce qui nous permet d'évaluer la représentation relative des genres. Par la suite, on aimerait affiner encore plus l'analyse et restreindre aussi au niveau des différents médias (TF1 FR2 FR3). Le GPI ajusté, en tant que mesure quantitative de la parité de genre, met en évidence les déséquilibres et nous oriente vers une analyse plus précise de la représentation des femmes et des hommes.

La formule pour calculer l'Indice de Parité de Genre ajusté (GPI ajusté) basé sur les comptes d'individus féminins et masculins dans nos données est la suivante :

$$\text{GPI ajusté} = \frac{\text{Nombre de Femmes} - \text{Nombre d'Hommes}}{\text{Nombre Total d'Individus}}$$

Dans cette formule, le *Nombre de Femmes* représente le total des individus identifiés comme féminins, le *Nombre d'Hommes* représente le total des individus identifiés comme masculins, et le *Nombre Total d'Individus* est la somme du *Nombre de Femmes* et du *Nombre d'Hommes*. L'identification du genre passe par une reconnaissance d'entité nommée et une fonction de gender prediction. Le résultat de cette formule donne un GPI ajusté qui varie entre -1 et 1. Les implications sont les suivantes :

- Un **GPI ajusté de 0** indique une parité parfaite entre les genres.
- Un **GPI ajusté positif** indique une prédominance de femmes.
- Un **GPI ajusté négatif** indique une prédominance d'hommes.

a. Visualisation par Cartes de Similarité et Heatmaps

Pour visualiser les résultats de la LDA et du GPI ajusté, on utilise des cartes de similarité et des heatmaps:

- **Cartes de Similarité** : Ces cartes positionnent les mots-clés et thèmes dans un espace bidimensionnel, où la proximité indique une similarité thématique. Les nuances de couleurs dénotent le genre associé, facilitant la compréhension de la distribution et de la représentation des genres.
- **Heatmaps** : On crée des heatmaps avec les mêmes données, y ajoutant les médias associés à chaque mot-clé ou thème. Ces cartes permettent de visualiser l'intensité de la représentation des genres et des thèmes dans divers médias, soulignant les zones de concentration et de disparité.

2. Création de Sous-Ensembles de Données et Clustering Hiérarchique

Basés sur les mots-clés extraits par la LDA, qui deviennent nos sous-ensembles de données, on applique un clustering hiérarchique. Cette analyse nous aide à explorer les relations entre les mots-clés, thèmes, genres, et médias de façon détaillée. Le dendrogramme produit illustre la proximité et les liens entre les groupes, du plus associé aux femmes au plus neutre, et permet d'identifier un top 5 des mots les plus associés aux femmes.

3. Approfondissement sur les Tonalités Émotionnelles et les Expressions Spécifiques

Analyse de Sentiment : On mène une analyse globale des sentiments sur tous les textes concernant les femmes, les classifiant en positif, neutre, ou négatif. Cela nous éclaire sur la représentation des femmes dans nos données et sur les potentiels biais.

Analyse des Expressions de Haine : En parallèle, on détecte et examine les expressions de haine, identifiant où les femmes sont ciblées par des discours haineux ou discriminatoires. L'objectif est de mettre en lumière les défis liés à la représentation des genres.

Enfin, pour la représentation des femmes parmi les présentateurs/réalisateurs et l'analyse des genres présentant des sujets de l'autre sexe :

- **Graphiques à Barres** : Pour montrer la proportion de femmes parmi les présentateurs/réalisateurs dans différents médias, offrant une comparaison visuelle directe.
- **Graphiques Circulaires ou en Donut** : Ces graphiques serviront à présenter la répartition des sujets liés au genre opposé présentés par des femmes et des hommes, donnant une vue d'ensemble de la dynamique de genre.

Mercredi

4.1 Application du Gender Parity Index

Dans le cadre de notre marathon du web, nous avons intégré le Gender Parity Index (GPI) à nos données. Ce travail a impliqué l'application de notre script à l'ensemble de notre corpus de données. Cependant, il est important de noter que l'application du GPI a révélé une prédominance d'articles jugés négatifs envers l'équilibre des genres, mettant en lumière une prédominance masculine dans les corpus analysés. Cette tendance a été encore plus évidente lors de l'examen d'une carte de chaleur (heat map) qui combine le GPI avec les clusters issus de l'analyse LDA, soulignant les zones spécifiques où les déséquilibres de genre sont les plus prononcés. Cette visualisation nous a permis de mieux comprendre la distribution des disparités de genre et d'identifier clairement les domaines nécessitant une attention particulière et une amélioration.

4.2 Développement du Site Web

Voici les pages clés :

- **Accueil** : Introduction de notre projet et aperçu des objectifs visés.
- **Analyse** : Détails sur l'application du Gender Parity Index, du Clustering LDA et autres visualisations qui en découleront.
- **Bibliographie** : Compilation de toutes les références et sources utilisées.
- **À propos** : Informations sur les créateurs du site et les motivations derrière le projet.

Nous avons également développé un footer complet pour faciliter la navigation sur le site et améliorer l'expérience utilisateur.

4.3 Analyse de Sentiment

Nous avons débuté une analyse de sentiment pour évaluer les perceptions et attitudes envers les femmes sur les clusters d'articles dis féminins, en utilisant des techniques avancées de traitement du langage naturel. Or, avec les résultats obtenus lors du gpi, la pertinence de l'analyse de sentiment est remise en cause.

Conclusion

Mercredi, nous devons tâcher de finir l'interface web et de finaliser les dernières visualisations de sorte à avoir un fil conducteur cohérent dans notre analyse.