

Perspectiva : projet OpenSource de Data4Good pour l'accélération du traitement des consultations publiques

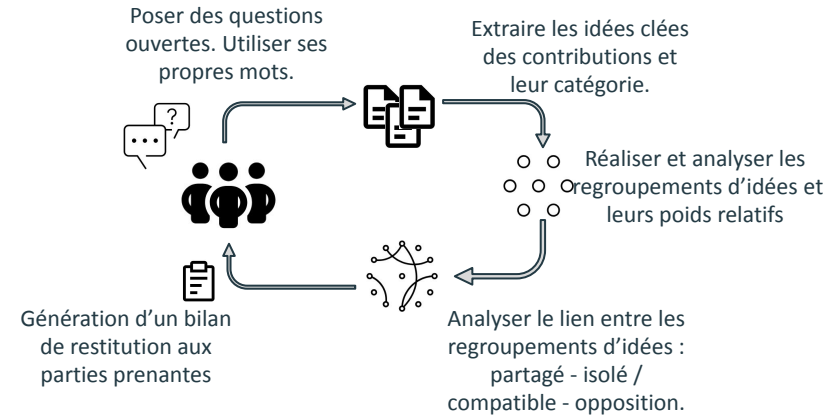
Objectif:

Analyser et comprendre ce qu'un large groupe de personnes pense en s'appuyant sur leurs propres mots, grâce aux outils intelligents du traitement du langage.

Principes :

- Des questions ouvertes pour que les contributeurs puissent s'exprimer plus librement avec leurs mots.
- Transparence et reproductibilité sur le traitement des propositions.
- Pas de catégorie à priori, donc réduction d'erreur de la grille d'analyse. Les regroupements se font sur la base de l'entière des contributions.
- Génération automatique de graphiques simples pour communiquer.
- Donner au collectif les moyens de se connaître
- Maintenir le dialogue avec les contributeurs par des cycles de traitement courts.
- Bilan adaptable selon besoins ou contraintes.

Etapes du traitement :



Type de données :

- Des milliers de contributions libres sur un sujet.
- Pour chaque contribution, les idées principales de cette contribution.
- Pour chaque idée, un vecteur "embeddings" est calculé.
- Les vecteurs sont regroupés par "cluster" à l'aide d'algorithmes de regroupement (ex: HDBSCAN).

Défi 1: représentation de la synthèse du traitement

Objectif

Avoir un schéma avec l'importance des idées, leur proximité et leur liens afin de pouvoir saisir en un coup d'oeil le réseau d'idée émergent.

Pourquoi

Une idée vient rarement seule. Elle s'inscrit souvent dans un réseau d'idée dont la caractérisation est un résultat en soit. Caractériser et représenter le lien entre les idées permet d'évaluer par exemple si deux idées majoritaires sont associées l'une à l'autre, si chacune s'inscrit dans un réseau d'idées distincts.

Exemple de signature

- Entrée :
 - N° du contenu d'origine
 - Ensemble d'idées représentées par des vecteurs
 - N° et Titre du regroupement d'idées auxquels sont rattachées les idées.
- Sortie :
 - Un graphique 2D des regroupements d'idées.
 - Les coordonnées du centre des regroupements.
 - La taille des regroupement.
 - La taille des liens qui relient les regroupements.

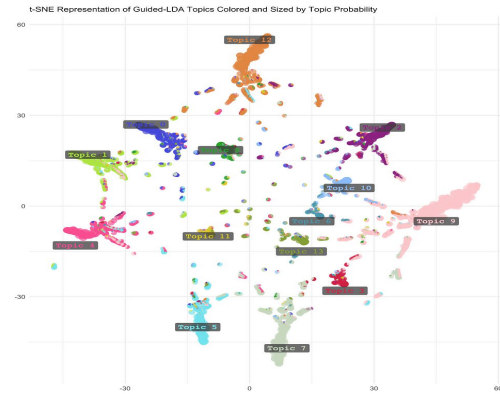


Figure 1 : 1 point = une idée. Difficulté de représenter l'importance d'une idée ou ses liens.

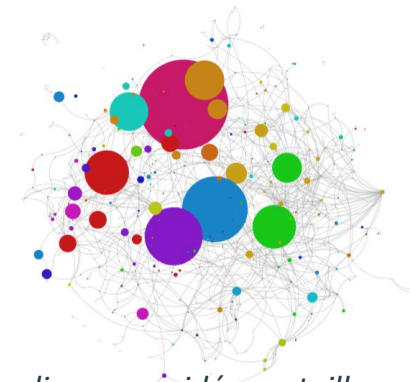


Figure 2 : Un disque une idée, sa taille son importance, et l'épaisseur du lien, la probabilité d'occurrence simultanée.

Défi 2: Création de persona

Objectif

Créer des personas afin de pouvoir se représenter les profils types des contributeurs.

Pourquoi

Les personas sont un outil de représentation pour créer un scénario réaliste. Il n'est pas un individu, ni la moyenne des individus, il est un motif qui semble fréquent dans l'ensemble de contribution.

Exemple de signature

- Entrée :
 - Ensemble des idées extraites pour chaque contributions représentées par des vecteurs.
 - Vecteur sémantique pour chaque idée.
 - N° et Titre du regroupement d'idées auxquels sont rattachées les idées.
- Sortie :
 - Nombre de personas
 - Par persona : % d'idée qui le représente.

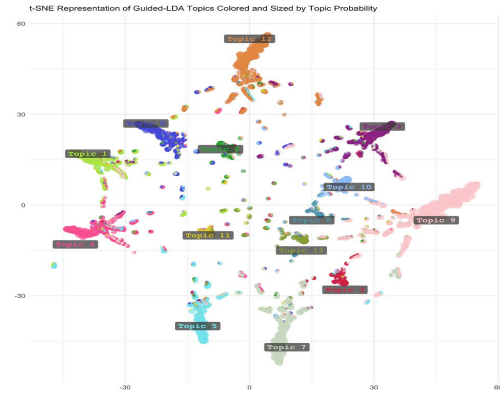


Figure 1 : 1 point = une idée. La distance entre les idées représentent leur similarité.

Persona 1 :
idée 23 (50%)
Idée 5 (5%)



Persona 2 :
idée 3 (46%)
Idée 10 (10%)

