

Quand l'Archelec rencontre l'IA

Génération automatique de métadonnées à partir de professions de
foi électorales françaises de 2017 à 2024

Projet 13 du hackaton organisé par l'École nationale des chartes la semaine du 5 janvier
2026

Objectif

Constituer la chaîne de traitement pour analyser automatiquement un lot de professions de foi d'élections législatives de 2017 à 2024 afin de générer les métadonnées alignées sur les référentiels fournis.

19 194 professions de
foi au format PDF.

Papier 100% recyclé Elections législatives juin 2017 Papier 100% recyclé

Alliance écologiste indépendante

Des citoyens en politique au-delà du clivage droite-gauche
Parmi les premiers mouvements politiques français

ILS VOUS PROMETTENT LA LUNE NOS 1000 CANDIDATS ET SUPPLÉANTS VOUS GARANTISSENT LA TERRE

**612 789 personnes dont Albert Jacquard
ont déjà choisi notre projet de société**

ÉCOLOGIE = EMPLOIS ET ACTIVITÉS POUR TOUS

Le chômage était inexistant voilà 50 ans, mais la population active agricole était de 15 % (3 % aujourd'hui). Le plein emploi est possible en stimulant les emplois environnementaux, ceux d'accompagnement, les AMAP, favoriser les petites exploitations, recenser les terres abandonnées pour les octroyer aux volontaires et avec une Économie locale : la paysannerie, les circuits courts et les monnaies locales complémentaires **donneront un sens et créeront des emplois indélocalisables**. Stopper les grands projets (LGV, aéroport, centres commerciaux...), favoriser une multitude de micro-projets, isoler TOUTES les constructions, transformer TOUTS les déchets en matières premières pour les réutiliser, développer le secteur de la réparation, est formidablement créateur d'emplois. **Chômage, impôts, déficits, massifs, sont créés par une Économie qui veut vendre toujours plus.**

La pénurie de ressources se fait sentir, les souffrances se multiplient, le "modèle" consumériste n'est pas viable. Faire réparer son appareil ou ses chaussures près de chez soi, c'est mieux que d'acheter un produit fabriqué à l'autre bout du monde...

"L'éducation au savoir être est le socle de notre projet de société"
Marie-Martine HULOT chargée de notre programme Éducation



Respecter l'humain, respecter l'animal

UNION POPULAIRE RÉPUBLICAINE
L'union du peuple pour rétablir la démocratie

ÉLECTIONS LÉGISLATIVES - 11 ET 18 JUIN 2017

Avec François ASSELINEAU Président de l'Union populaire républicaine



L'UPR : le seul vote vraiment nouveau et utile

Résultat d'une gigantesque opération de propagande, l'éllection d'Emmanuel Macron laisse la scène politique en ruines. Tous les anciens partis sont au bord de l'implosion, tandis que « La République en marche » recrute de façon précipitée et brouillonne des politiciens opportunistes de tout poil.

Le point commun de tous ces dirigeants et de leurs partis est de faire semblant. Ils vous parlent comme si la France était toujours dirigée en toute liberté par nos élus. Ils vous cachent la relation de cause à effet qui existe entre notre appartenance à l'Union européenne et les politiques d'austérité et de régression sociale qui nous sont imposées quels que soient les dirigeants élus. Ils vous dissimulent l'essentiel : les traités européens sont contraignants et la France est soumise aux décisions de Bruxelles.

Au cours de la campagne présidentielle, François ASSELINEAU, président-fondateur de l'UPR, a été le seul à dénoncer le rapport annuel de la Commission européenne sur les « grandes orientations des politiques économiques » (GODEP). Un document permet à chacun de dissiper le flou du programme de M. Macron. C'est ce rapport qui « recommande » les mesures qui seraient appliquées si « La République en Marche » était majoritaire à l'Assemblée : l'ubérisation généralisée de la société, la fin des professions réglementées, la baisse du SMIC, des prestations chômage et des retraites, l'augmentation de la TVA, la légitimation de l'évasion fiscale, le démantèlement du droit du travail, le braquage des industries stratégiques françaises, les fusions forcées de communes, la disparition programmée des départements...

Tous les autres partis savent que leurs finis par exploser, mais ils n'en parlent pas. Cette explosion ravagera nos économies, alors que le retour planifié au franc ne causera pas plus de difficultés que n'en avait créé le fameux « bug de l'an 2000 ».

Face à ces silences, l'Union populaire républicaine (UPR) apporte un formidable renouveau, en disqualifiant tous les partis installés, libérés et chorégraphiés. Car l'UPR a fait, une fois pour toutes, le choix de l'honnêteté et du courage, en proposant aux Français de récupérer leur indépendance nationale.

Seule notre souveraineté retrouvée permettra de rendre aux Français leur démocratie réelle, de protéger leur emploi, leur niveau de vie, leur sécurité et leur justice sociale, et de restaurer la France comme puissance libre et pacifique, sans démagogie raciste ou populisme xénophobe.

upr.fr

Elections législatives - 11 & 18 juin 2017

LA FORCE DU PEUPLE



**Pour une majorité insoumise
à l'Assemblée Nationale**

**“ Je vous demande de voter pour Sonia Naffati comme vous avez voté pour moi.
À l'Assemblée nationale, nous serons vos députés,
insoumis aux puissants et fidèles aux intérêts du peuple.”**

François Mitterrand

SONIA NAFFATI

est engagée à mes côtés avec courage.
Elle a ma confiance.

SUPPLÉANT
EMMANUELLE SOMER
Musicienne - agricultrice citoyenne solidaire

Dans la 2e circonscription des Alpes-Maritimes,
votez pour les candidats de

La France insoumise
www.lafranceinsoumise.fr

JULIETTE CHESNEL-LE ROUX
titulaire
Conseillère municipale écologiste à Nice,
engagée depuis 25 ans dans l'économie sociale
et solidaire et dans la défense de
l'environnement, ingénieur, mère de famille.

FABRICE DECOUIGNY
remplaçant
Conseiller municipal et métropolitain
écologiste à Nice, engagé dans les politiques
de proximité et les problématiques urbaines,
enseignant-chercheur en géographie à
l'Université de Nice, père de famille.

@juliette.chesnel.3
elussevel.denic@yaho.fr

Madame, Monsieur,

L'écologie a été la grande oubliée de la présidentielle. Pour que l'environnement, la préservation de la biodiversité et la transition énergétique soient portés lors des 5 prochaines années, il faudra des écologistes à l'Assemblée.

Nous comptons sur vous !

Une société écologiste, c'est une société qui protège le climat, en sortant des énergies fossiles et du nucléaire. Une société où vous mangez des produits sains, issus de l'agriculture bio et locale, où l'air et l'eau ne sont pas pollués. C'est une société qui crée un million d'emplois dans la transition écologique tout en réduisant vos factures d'énergie. Une société solidaire et ouverte, avec des services publics partout sur le territoire, pour vous soigner, pour vous protéger, pour éduquer nos enfants. Une société qui propose un toit et des revenus à chacune et chacun d'entre nous, qui n'a pas peur de l'autre et qui accueille les personnes en détresse. C'est une société qui construit un monde de paix et l'Europe de demain.

Pourons ensemble le projet d'une France désirable pour notre environnement, notre santé et nos enfants.

JULIETTE CHESNEL-LE ROUX

Législatives des 11 et 18 juin 2017
1^{ère} circonscription des Alpes-Maritimes

Europe
écologie
les Verts

Exemples
de
profession
de foi

[illegible]

Référentiel

Travail préliminaire

- Suppression des documents _falc de la base de données parce qu'ils constituaient des doublons.
- La base des données a ainsi été réduite à 14 631 documents.

Partie 1 – Extraction de textes à partir des pdfs

- Récupération des textes des professions de foi avec deux méthodes :
 - PyMuPDF (fitz)
 - Mistral OCR
- Dans l'exemple à côté, ce qui est encadré en noir est le contenu à extraire.

UNION POPULAIRE RÉPUBLICAINE
L'union du peuple pour rétablir la démocratie

ÉLECTIONS LÉGISLATIVES – 11 ET 18 JUIN 2017
Avec François ASSELINEAU
Président de l'Union populaire républicaine

L'UPR : le seul vote vraiment nouveau et utile

Résultat d'une gigantesque opération de propagande, l'élection d'Emmanuel Macron laisse la scène politique en ruines. Tous les anciens partis sont au bord de l'implosion, tandis que « La République en marche » recrute de façon précipitée et brouillonne des politiciens opportunistes de tout poil.

Le point commun de tous ces dirigeants et de leurs partis est de faire semblant. Ils vous parlent comme si la France était toujours dirigée en toute liberté par nos élus. Ils vous cachent la relation de cause à effet qui existe entre notre appartenance à l'Union européenne et les politiques d'austérité et de régression sociale qui nous sont imposées quels que soient les dirigeants élus. Ils vous dissimulent l'essentiel : les traités européens sont contraignants et la France est soumise aux décisions de Bruxelles.

Au cours de la campagne présidentielle, François ASSELINEAU, président-fondateur de l'UPR, a été le seul à dénoncer le rapport annuel de la Commission européenne sur les « grandes orientations des politiques économiques » (GPE). Lire ce document permet à chacun de dissiper le flou du programme de M. Macron. C'est ce rapport qui « recommande » les mesures qui seraient appliquées si « La République en Marche » était majoritaire à l'Assemblée : l'ubérisation généralisée de la société, la fin des professions réglementées, la baisse du SMIC, des prestations chômage et des retraites, l'augmentation de la TVA, la légitimation de l'évasion fiscale, le démantèlement du droit du travail, le bradage des industries stratégiques françaises, les fusions forcées de communes, la disparition programmée des départements...

Tous les autres partis savent que l'euro finira par exploser, mais ils n'en parlent pas. Cette explosion ravagera nos économies, alors que le retour planifié au franc ne causera pas plus de difficultés que n'en avait créé le fameux « bug de l'an 2000 ».

Face à ces silences, l'Union populaire républicaine (UPR) apporte un formidable renouveau, en disqualifiant tous les partis installés, lâches et discrédités. Car l'UPR a fait, une fois pour toutes, le choix de l'honnêteté et du courage, en proposant aux Français de récupérer leur indépendance nationale.

Seule notre souveraineté retrouvée permettra de rendre aux Français leur démocratie réelle, de protéger leur emploi, leur niveau de vie, leur sécurité et leur justice sociale, et de restaurer la France comme puissance libre et pacifique, sans démagogie raciste ou populisme xénophobe.

upr.fr

Partie 1 – Extraction de textes à partir des pdfs

filename	page_1	page_2
LG17-1-1-BONNOT-10-tour1-profession_foi.pdf	Vincent GUERIN Gilbert BONNOT ...	Ce que nous voulons ...

Méthode 1 – Bibliothèque
PyMuPDF (fitz)

Partie 1 – Extraction de textes à partir des pdfs

filename	page_1	page_2
LG17-1-1-BLATRIX-CONTAT-2-tour1-profession_foi.pdf	Florence BLATRIX-CONTAT — présentation personnelle et parcours politique...	Programme législatif et engagements (gauche, justice sociale, écologie, services publics)...

Méthode 2 – Mistral OCR

Partie 2 – Prédiction et classification des informations par appel à un LLM

4	colonne	exemple	commentaire
5	dc:identifier	LG17-75-1-VÉRON-12-tour1-profession_foi.pdf	nom du fichier dans Regards citoyens. Ce nom, qui peut varier selon les années, contient plusieurs infos demandées en métadonnées : LG = législatives 17 = 2017 75 = code du département 1 = circonscription VERON = nom du ou de la candidat(e) titulaire 12 = ? numéro séquentiel propre à RC ? tour1 = tour Ces informations se trouvent aussi sur le document même, mais n'hésitez pas à utiliser le nom du fichier pour générer automatiquement les métadonnées correspondantes.

Extrait du référentiel utilisé pour
écrire le prompt

Partie 2 – Prédiction et classification des informations par appel à un LLM

```
1      """
2          Tu es chercheur en sciences politiques. Ta tâche est d'associer des informations contenues des textes issus de fichiers csv à
3
4          (dc:identifiant | LG17-75-1-VÉRON-12-tour1-profession_foi.pdf | nom du fichier dans Regards citoyens.
5          Ce nom, qui peut varier selon les années, contient plusieurs infos demandées en métadonnées :
6          LG = législatives
7          17 = 2017
8          75 = code du département
9          1 = circonscription
10         VERON = nom du ou de la candidat (e) titulaire
11         12 = ? numéro séquentiel propre à RC ?
12         tour1 = tour
13         Ces informations se trouvent aussi sur le document même, mais n'hésitez pas à utiliser le nom du fichier pour générer automati
14
15         (ba_nom_titulaire | Véron | la particule, s'il y en a une, est à mettre à la suite du prénom, pas avant le nom. Si tu ne trouv
16
17         (bb_prenom_titulaire | Pauline | les prénoms composés comportent systématiquement un trait d'union. Si tu ne trouves pas de pr
18
19         (bd_age_titulaire | 43 | soit l'âge, soit l'année de naissance, selon ce qui est indiqué dans le document.),
20
21         (be_professionn_titulaire | juriste | attention: on demande la profession actuelle des candidats, pas le métier qu'ils ont pu
22
23         (bj_soutien_titulaire | Parti socialiste / Union des démocrates et des écologistes / Génération écologie / Parti radical de ga
24
25         (bk_liste_titulaire | Ensemble une force pour Paris | appellation temporaire à l'occasion de l'élection ex: NUPES, éventuellem
26
27         (ca_nom_suppléant | Eustache | la particule, s'il y en a une, est à mettre à la suite du prénom, pas avant le nom.),
28
29         (cb_prenom_suppléant | Guillaume |),
30
31         (cd_age_suppléant | |),
32
```

Prompt final

Partie 2.2 – Test du prompt

fichier_echantillon_5_test_mistral		
filename	page_1	page_2
LG17-1-1-BLATRIX-CONTAT-2-tour1-profession_foi.pdf	Michel FONTAINE	NATIONALEMENT, je serai une députée constructive mais exigeante
	Florence Votre Députée suppléant	vis-à-vis des réformes proposées par le président de la République,
	BLATRIX-CONTAT	fidèle aux valeurs qui fondent mon engagement politique à Gauche.
	A gauche	Je soutiendrai
	pour faire	→ La moralisation de la vie publique et l'interdiction du cumul de plus de
	réussir la France	trois mandats successifs pour permettre le renouvellement
	51 ans,	Michel FONTAINE
	mariée, 3 enfants	→ Les projets visant à concilier efficacité économique et justice sociale ainsi
	Madame, Monsieur,	Premier-adjoint de la Ville qu'une action forte en matière de formation pour lutter contre le chômage
	Le 7 mai dernier, le peuple français a fait le choix d'un nouveau	de Bourg-en-Bresse
	Je vis dans mon village natal à	→ La priorité donnée à l'école primaire pour que chaque enfant puisse réussir
	président de la République.	Président de l'Aggloméra-
	Drom, au cœur du Revermont.	et le soutien à l'autonomie des jeunes par une revalorisation des bourses
	Les élections législatives des 11 et 18 juin décideront de l'avenir de	tion de 2008 à fin 2016
	Mes grands-parents agricul- la France ; elles seront l'occasion de choisir la République que	et depuis Vice-président
	teurs et mes parents m'ont voulons. Je me présente à vos suffrages avec la volonté de faire	→ Les mesures en faveur du pouvoir d'achat pour les plus modestes avec
	transmis leur attachement à sir la France, de faire réussir ce quinquennat. La majorité issue c	au développement
	ces terres ; ils m'ont surtout scrutin devra agir pour la justice sociale, la solidarité, l'égalité de	l'augmentation du minimum vieillesse et de la prime d'activité
	appris que l'école était un chances et la transition écologique. Pour cela, la gauche doit être	économique.
	moyen d'émancipation et à l'Assemblée nationale.	Je vis sur Bourg-en-Bresse → La transition écologique et énergétique : des investissements dans la
	d'ascension sociale.	depuis plus de 40 ans, croissance verte et les économies d'énergie, l'interdiction des perturbateurs
	Demain, à l'Assemblée, je veux représenter une gauche constructive et	j'ai partagé ma vie endocriniers

Extrait du fichier-échantillon

Partie 2.2 – Test du prompt


dc:identifiant	ba_nom_titulaire	bb_prenom_titulaire	bd_age_titulaire	be_professionn_titulaire
LG17-1-1-BLATRIX-CONTAT-2-tour1-profession_foi.pdf	BLATRIX-CONTAT	Florence	51	enseignante
LG17-1-1-BONNOT-10-tour1-profession_foi.pdf	BONNOT	Gilbert	NR	NR
LG17-1-1-BRETON-5-tour1-profession_foi.pdf	BRETON	Xavier	54	député / manquant
LG17-1-1-BRETON-5-tour2-profession_foi.pdf	BRETON	Xavier	54	député / manquant
LG17-1-1-CARLIER-6-tour1-profession_foi.pdf	ASSELINEAU	François	NR	NR

Extrait du fichier obtenu avec
erreurs annotées en rouge

dc:identifiant	ba_nom_titulaire	bb_prenom_titulaire	bd_age_titulaire	be_professionn_titulaire	bj_soutien_titulaire
LG17-1-1-BLATRIX-CONTAT-2-tour1-profession_foi.pdf	BLATRIX-CONTAT	Florence	51	enseignante	Parti socialiste / PRGR / Europe Écologie Les Verts / UDE
LG17-1-1-BONNOT-10-tour1-profession_foi.pdf	BONNOT	Gilbert	NR	Moniteur éducateur	NR
LG17-1-1-BRETON-5-tour1-profession_foi.pdf	BRETON	Xavier	54	député / conseiller régional	Les Républicains / UDI / CPNT
LG17-1-1-BRETON-5-tour2-profession_foi.pdf	BRETON	Xavier	54	député / conseiller régional	Les Républicains / UDI / CPNT
LG17-1-1-CARLIER-6-tour1-profession_foi.pdf	NR	NR	NR	NR	Union populaire républicaine


Extrait du fichier obtenu avec
erreurs corrigées en vert

Partie 2.2 – Test du prompt



Mistral Large 3
A state-of-the-art, open-weight, general-...
v25.12

OPEN



Ministral 3 3B
A tiny and efficient model offering best-in-cla...
v25.12

OPEN

Modèles utilisés

Partie 3 – Test final

```

~ OCR FITZ / PuymyPDF

[12] batch_test_2 = pd.read_csv("/Users/Manon/Desktop/Fac/ENC/Hackaton/SciencesPo/Data_OCR/pdf_pages_extracted_final.csv")
Python

[13] batch_test_2.head()
Python
...

[14] batch_test_fitz = batch_test_2.sample(n=550)
Python

[15] batch_test_fitz
Python
...

[17] for i in range(50):
    batch_test_2 = batch_test_fitz.iloc[i * 11 : (i + 1) * 11]
    assert(len(batch_test_2)==11)
    batch_test_2.to_csv("batch_test_fitz_" + str(i), index=False)
Python
```

Division des données en
batches

Partie 3 – Test final

Chez Mistral	Chez OpenAI
Mistral Large 3	ChatGPT
Mistral Small 3.1	GPT-5-nano
Ministral 3 3B	

Modèles utilisés

Partie 3 – Test final

	Bacth0_OCR_Mistral	Bacth0_OCR_fitz
Mistral_Large_3	Différences trouvées: 106/286 cellules (37.06%) Colonnes manquantes: 5	Différences trouvées: 141/341 cellules (41.35%) Colonnes manquantes: 0
Mistral_Small	Différences trouvées: 64/341 cellules (18.77%) Colonnes manquantes: 0	Différences trouvées: 43/341 cellules (12.61%) Colonnes manquantes: 0
Ministral_3B	Différences trouvées: 133/319 cellules (41.69%) Colonnes manquantes: 2	Différences trouvées: 210/275 cellules (76.36%) Colonnes manquantes: 6
ChatGPT	Différences trouvées: 118/341 cellules (34.60%) Colonnes manquantes: 0	NR

Résultats

Partie 3 – Test final

Modèle	Accuracy globale
Mistral_Large_3	60.83%
Mistral_Small	84.31%
Ministral_3B	41.99%
ChatGPT	65.40%

Résultats