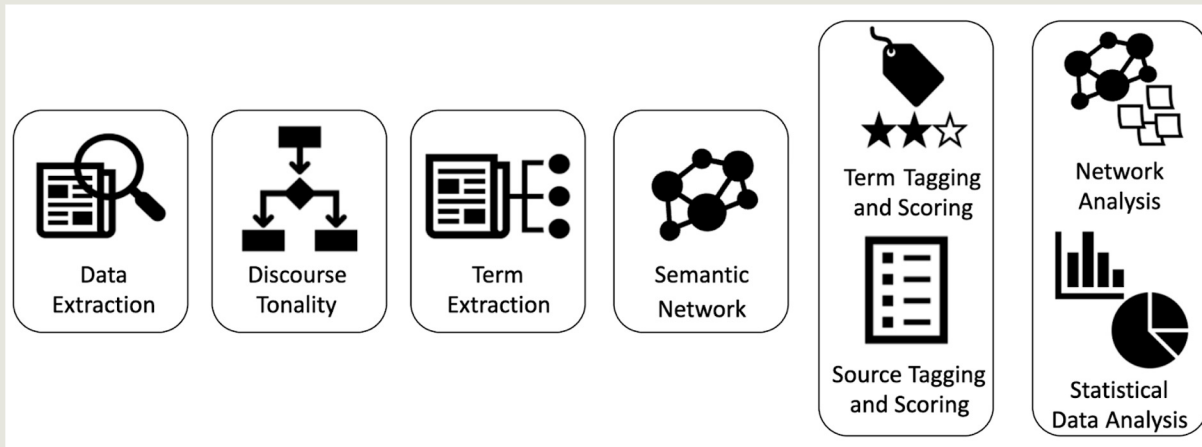


– Media Layer / Methodology Protocol –

Overview



45,608 articles / 281 sources

SEARCH dans le title and lead paragraph:

"intelligence artificielle" OR "IA" OR "algorithm*" OR "apprentissage profond" OR "apprentissage machine" OR "réseau* de neurone*" OR "machine learning" OR "deep learning" OR "neural network*"

DATES: 01.01.2011- 01.01.2021

MEDIA: Origin France -

Europresse.com

LANGUAGE: French

Manual Annotation: 1,977 articles

Double blind + third party control for problematic cases

fastText Model Tonality Detection

precision 0.79

recall 0.79

f1-score 0.79

Semantic Network Analysis

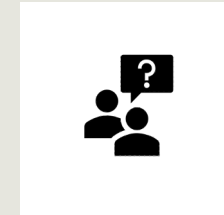
6,802 terms (mono/bi/tri-gram)

75,342 links

29 clusters / 11 meta-clusters

80% entities categorized

10 tags : technical, people, issue, company, etc.



5 interviews with journalists

Integrated in the P2 protocol

M 31 / Political press news media

M 71 / Technology and Science Website

F 45 / Mainstream news media / Technology

F 31 / Mainstream news media / Technology

M 50 / Mainstream news media and radio / Technology

Data Extraction

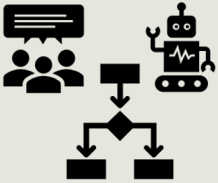
Data Extraction - Europress



- QUERY : "intelligence artificielle" OR "IA" OR "algorithm*" OR "apprentissage profond" OR "apprentissage machine" OR "réseau* de neurone*" OR "machine learning" OR "deep learning" OR "neural network*"
- IN TITLE or LEAD of the article
- DATES: 01.01.2011- 01.01.2021:
- MEDIA: Origin France, language French

FULL DATASET = 45608 articles (without duplicates)

Discourse Tonality - fasText



- CATEGORIES : Collaborative Definition and Guidelines - “Neutral” / “Critique” / “Promise”
- TAGGING : 4 coders - Double Blind - 2000 articles (Title + Lead)
- FASTEXT - 2 rounds (optimise detection of critique / Promise)
- MANUAL EVALUATION - 200 articles FScore = 0.71 (see annexe 1)
- DATASET NEUTRAL= 32447 articles

CRITIQUE (4261) vs PROMISE (8900) DATASET = 13161 articles

Term Extraction - Cortext

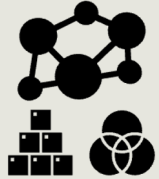


- DATASET CRITIQUE (4261) vs PROMISE (8900) = 13161 articles
- TERM EXTRACTION in Title + Content
8000 Terms Monogram/Bigram/Trigram)
Pigeon_Hole method - Sentence Level- Noun_Phrase - Min Frequency 3
- MANUAL CLEANING (duplicates, numbers, journal, dates, errors, noise)
- CORPUS INDEXATION - corpus list indexer

TERM DATASET = 7259 terms

Visualization / Tagging & Scoring

Semantic Network - Cortext / Gephi



- NETWORK : 6802 nodes / 75342 links of cooccurrences
- VISUALIZATION : Spatialization : Force_atlas_2 / Node color : cluster_level_2 / Size : term frequency
- CLUSTERS level_1: 29 clusters (> to 4 articles) louvain algorithm
- CLUSTERS level_2 : 11 clusters (manual aggregation of clusters by topic and topological proximity)

Term Tagging and Scoring - Cortext



- TERM MANUAL TAGGING : 80% term covered (20% non relevant)
10 tags : "Technical", "Company", "Fiction", "Person", "People", "Institution", "Locality", "Data", "Topic", "Event"
- TERM TONALITY SCORING : 93% terms covered (7% excluded from contrast analysis)
Raw Ratio = (freq prom / total freq) - (freq critiq / total freq)
Log Ratio = $\log((\text{freq crit}+1) / (\text{freq prom}+1))$
Normalization : $x'' = (b-a)(x-\min x) / (\max x - \min x) + a$, a=-1, b=1

Source Tagging and Scoring - Cortext



- SOURCE TAGGING
Recoded sources : 281 sources (702 raw list)
Diffusion : Nationale, Locale
Topic : Generaliste, Eco, Tech, Sante, Transport, Envir, Sport, Autre
- SOURCE SCORING : contrast analysis

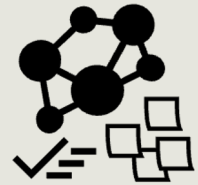
Statistical & Network Analysis

Statistical Data Analysis - Cortext - Rawgraph



- Corpus, cluster and tonality Temporal analysis
- Contingency Matrix (sources, clusters, tonality)
- Contrast analysis
- Verb, NER extraction and analysis

Network Analysis - Gephi



- Topological analysis (polarity, center and periphery, tonality)
- Cluster internal analysis and comparison (entities and topics)
- Entity Layers comparison

Interviews

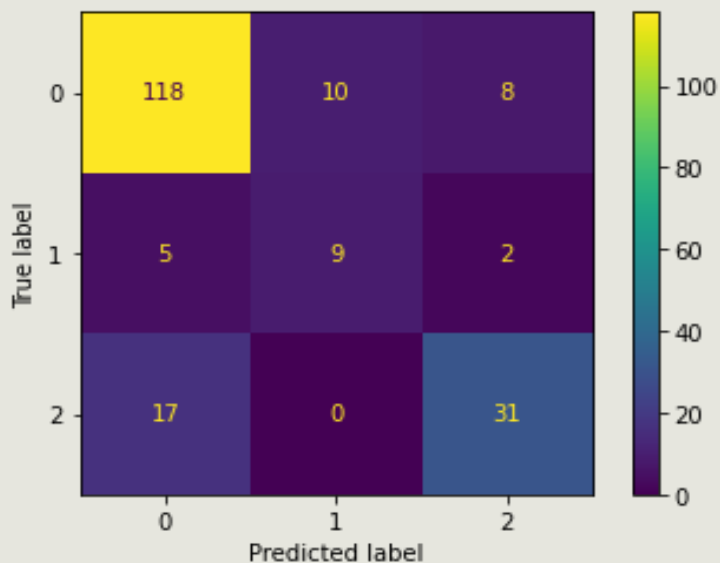


Interviews with journalists

- 5 interviews with journalists
- Interview guideline (annexe 2)
- Integrated in the P2 protocol

Gender	Age	Education	Professional Activity
M	31	Journalism	Political press news media
M	71	Publishing	Technology and Science Website
F	45	Journalism	Mainstream news media / Technology
F	31	Journalism	Mainstream news media / Technology
M	50	Letters and Humanities	Mainstream news media and radio / Technology

Annexe 1 – Fastext model evaluation



	precision	recall	f1-score		support
0		0.84	0.87	0.86	136
1		0.47	0.56	0.51	16
2		0.76	0.65	0.70	48
accuracy					0.79
macro avg	0.69	0.69	0.69		200
weighted avg	0.79	0.79	0.79		200

Annexe 2 – Guide d'entretien

- Pourriez-vous me raconter votre parcours professionnel en tant que journaliste et la manière dont vous vous êtes spécialisé sur les questions liées au numérique ? (formation, découverte, médias, évolution, étapes clés carrières, ...)
- Pourriez-vous me raconter le premier article que vous avez rédigé/publié sur l'IA ? (sujet, média, déclencheur, année, tonalité, maîtrise des aspects techniques, enjeux, etc..)
- Qu'est ce qui a évolué depuis selon vous ? sur la technologie en elle-même ? mais aussi sur la manière dont vous traitez le sujet aujourd'hui ? (fréquence de publication et volume d'articles, complexification techniques, médiatisation des enjeux , eco, éthique, etc..)
- Aujourd'hui, quels sont les enjeux autour de ces technologies que vous mettez en avant dans vos publications ? Pour quelles raisons elles vous paraissent importantes ? (sujets développés, ligne éditoriale, exemple, argumentaires déployés, tonalité, détails techniques ou vulgarisation, public ciblé)
- Quels ont été les sujets les plus clivants/controversés que vous avez pu traiter dans vos articles sur l'IA ? Pour quelles raisons vous paraissent-ils importants ?
- De manière plus générale, comment selon vous les médias français couvrent les sujets liés à l'IA ? (thèmes, tonalités, critique et enjeux, promesses et développement, lacunes de la couverture médiatique, comparaison française et internationale...)
- Selon vous, quels acteurs ont eu une influence sur le développement de l'IA en France ? (économique, politique, recherche, ong, etc.) Quel rôle ont-ils joué ?
- Pourriez-vous me raconter en me donnant quelques exemples, Quelles sont les avancées majeures qui ont permis le développement de ce secteur en France ? (événement marquant, décisions politiques et réglementations, financement, innovation technique)
- Comparé à d'autres pays, quelles sont les particularités françaises ? (techniques, eco, politiques, avantages et limites du développement local de ces technologies)
- Quels sont les sujets que vous aimeriez développer davantage sur ce thème aujourd'hui ? (couvert par d'autres médias, pas assez développé, causes, ...)