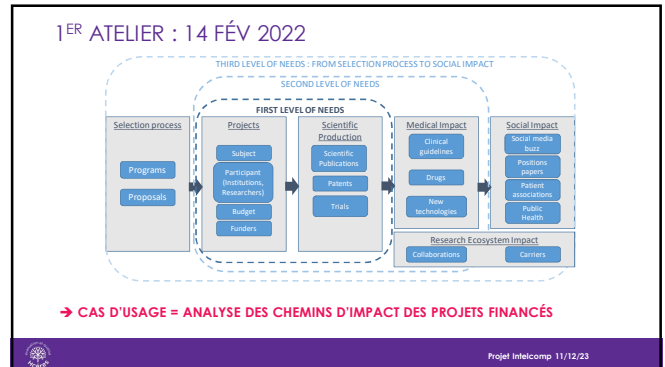




Projet européen INTELCOMP

LL SANTÉ
PRÉSENTATION AU GT FINANCEURS
11/12/2023

ORDRE DU JOUR

1. RAPPEL DU CONTEXTE ET DÉROULEMENT DU PROJET
2. PRÉSENTATION DES EXPLORATIONS RÉALISÉES : CONSTITUTION DES CORPUS & EXPLORATIONS THÉMATIQUES

Projet Intelcomp 11/12/23

OUTILS DÉVELOPPÉS DANS LE CADRE D'INTELCOMP

STI Viewer

- for policy makers & public administration
- to visualize (un)structured data
- to analyse datasets using NLP pipelines, machine translation, timelines

<https://sti-viewer.intelcomp.bsc.es/home>

Interactive model trainer

- for technical & advanced users
- to train their own (topic or classification) models

STI participation portal

- for academic & industry stakeholders, citizens
- to visualize the STI panorama & its evolution, the log between the STI outcomes and the social media impact
- to share views and feedback through Participation mailbox

Evaluation workbench

- for public administration & funding entities
- to assist in the evaluation process of STI proposals; identify evaluators; contextualize the proposal with existing patents, publications, projects; classify proposals automatically; check similar proposal

Projet Intelcomp 11/12/23

INTELCOMP : CONTEXTE & OBJECTIF

PROJET EUROPÉEN SUR 3 ANS (2021-2023)

PRINCIPAUX ACTEURS

■ FEICYT	■ OPENAIRE
■ UC3M	■ ARC
■ BSC	■ CITE
■ MINECO- SEDIA	■ HFRI
■ NIT DATA	
■ TILDE	■ OST - HCERES
■ TG	■ ZSI

OBJECTIFS

- Construire une plateforme de connaissance (données et outils d'IA) permettant de proposer des services d'exploration et de visualisation de données dans le domaine de la recherche et de l'innovation.
- Cible principale : **administrations publiques et acteurs des politiques de R&I** ; les services envisagés doivent aussi impliquer d'autres parties prenantes (chercheurs, industriels, citoyens...)
- Développer la plateforme en co-construction avec les utilisateurs grâce à des Living Lab : Changement Climatique, Intelligence Artificielle, Santé (pilote par l'OST)

Projet Intelcomp 11/12/23

2^{ÈME} ATELIER : 26 JUIN 2023

EXPLORATIONS RÉALISÉES

- Données projets et publications disponibles pour l'INCa
- Exploration des outils d'IA pour identifier les publications issues des projets financés par l'INCa

PRINCIPALES CONCLUSIONS

- IA n'est pas adaptée à tous les types de projets
- IA doit être combinée à d'autres variables (auteurs, structures, dates) + échantillon de publications de contrôle
- Ce n'est pas le cœur du projet INTELCOMP

SUITE DU PROJET

- Financeurs avec des publications identifiées dans Intelcomp : CE, NIH
- Test des outils INTELCOMP sur le cas d'usage des projets financés par la CE

Projet Intelcomp 11/12/23

3^{ÈME} ATELIER : 11 DÉCEMBRE 2023

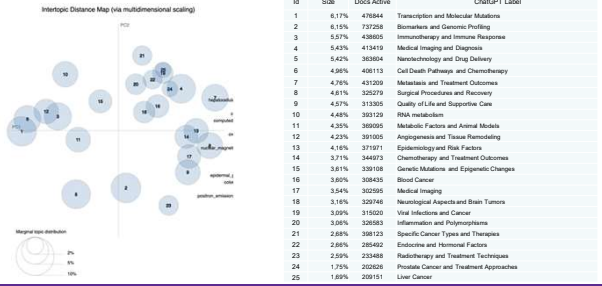
EXPLORATIONS RÉALISÉES

- Intégration du corpus de publications sur le cancer construit avec l'INCa en mobilisant des experts du cancer
- Constitution des corpus de brevets et de projets européens sur le cancer en mobilisant les outils de classification
- Exploration thématique des publications sur le cancer
- Comparaison du positionnement thématique des publications Monde - Europe - France et Publications financés par l'UE



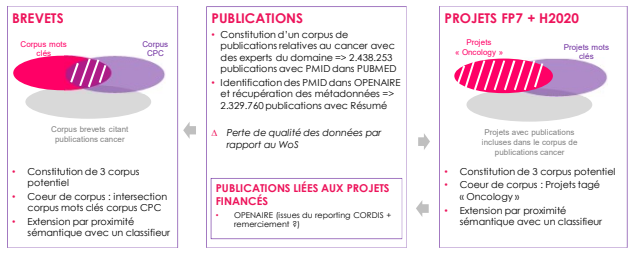
Projet Intelcomp 11/12/23

TOPIC MODELING DU CORPUS DE PUBLICATIONS



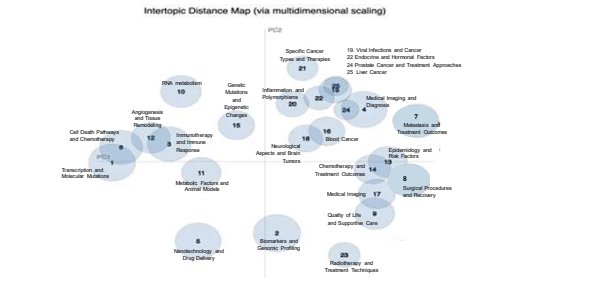
Projet Intelcomp 11/12/23

CONSTITUTION DES CORPUS RELATIFS AU CANCER



Projet Intelcomp 11/12/23

TOPIC MODELING DU CORPUS DE PUBLICATIONS



Projet Intelcomp 11/12/23

EXPLORATION THÉMATIQUE PAR UN TOPIC MODELING RÉALISÉ PAR UC3M

TOPIC MODELING RÉALISÉ SANS EXPERTS FAUTE DE TEMPS

POINT CLÉS DE LA MÉTHODE MISE EN ŒUVRE PAR UC3M

- Constitution du corpus à partir des PMID transmis par l'OST + rapprochement des données OpenAire
- Traitement TF IDF + Intégration de 350 STOP WORDS dont de nombreux stop words considéré comme générique mais qui nous semble « méthodo »
- Choix d'un TM en 25 topics et non 40 topic
- Labellisation des topic avec ChatGPT

RÉSULTATS

- Perte de 30% du corpus
- Topics qui semblent plus ciblés mécanismes de biologie fondamentale ou pathologies... faible représentation de topic méthodologique de recherche clinique, prévention, diagnostic



Projet Intelcomp 11/12/23

ID	Size	Docs Active	ChatGPT Label	Word description
1	6.17%	476844	Transcription and Molecular Mutations	protein, binding, transcription, DNA, kinase, promoter, mutation, phosphorylation, mouse, cellular, myc, oncogene, suppressor, ribonucleic, acid, ras, mutations, messenger, rna
2	6.15%	737258	Biomarkers and Genomic Profiling	molecule, biomarker, biological, therapeutic, genome, signature, classification, cellular, immunity, metabolic, heterogeneity, high, throughput, profiling, proteomic, immunotherapy
3	5.57%	438605	Immunotherapy and Immune Response	immunity, mouse, dendritic, cell, cytokine, lymphocyte, inflammation, macrophage, tumor, necrosis, factor, antigen, vaccine, nk, cells, lipopolysaccharide, antibody, immunotherapy, necrosis, factor
4	5.43%	413419	Medical Imaging and Diagnosis	cystic, benign, metastasis, dog, imaging, neoplasm, surgery, mass, renal, pancreatic, computed, tomography, soft, tissue, malignant, differential, diagnosis, child
5	5.42%	363604	Nanotechnology and Drug Delivery	nanoparticle, cytotoxicity, binding, peptide, synthesis, photodynamic, therapy, vno, downcastin, conjugate, anticancer, DNA, molecule, neuromuscular, fluorescence, imaging
6	4.96%	406113	Cell Death Pathways and Chemotherapy	apoptosis, protein, kinase, nf, kappaB, protein, kinase, h, autophagy, phosphorylation, mitochondrial, death, cytosolic, caspase, casp8in, chemotherapy, mapk, bax, activation, vno
7	4.76%	431209	Metastasis and Treatment Outcomes	metastasis, prognosis, overall, survival, lymph, node, surgery, recurrence, resection, chemotherapy, heart, rate, bladder, adjuvant, radiotherapy, preoperative, multivariate, hazard, ratio
8	4.61%	325279	Surgical Procedures and Recovery	surgery, resection, postoperative, reconstruction, laparoscopic, endoscopic, flap, surgery, preoperative, intraoperative, excision, peroperative, pain, rectal, defect, hospital, stay
9	4.57%	313305	Quality of Life and Supportive Care	quality, life, pain, survivor, child, breast, exercise, fatigue, palliative, care, physical, interview, training, life, nurse, social, psychosocial
10	4.48%	393129	RNA metabolism	micro, rna, protein, colorectal, cancer, tissue, messenger, rna, metastasis, gastric, overexpression, ribonucleic, acid, epithelial, mesenchymal, transitions, gas, chromatographic, pancreatic, nasopharyngeal, carcinoma, prognosis, apoptosis
11	4.35%	369695	Metabolic Factors and Animal Models	fat, mouse, metabolic, diet, antioxidant, inflammation, enzyme, animal, cytotoxicity, fatty, acid, DNA, oxidative, toxicity, specie, liver
12	4.23%	391005	Angiogenesis and Tissue Remodeling	vascular, endothelial, growth, factor, angiogenesis, stem, mouse, endothelial, metastasis, migration, extracellular, matrix, integrin, hypoxia, adhesion, matrix, metalloproteinase, inhibitors, tissue, vno, fibroblast
13	4.16%	371971	Epidemiology and Risk Factors	colorectal, cancer, mortality, smoker, confidence, interval, cohort, HIV, colonoscopy, death, heart, rate, lung, relative, risk, hazard, ratio, diet, poly, sex



Projet Intelcomp 11/12/23

#	Size Active	Docs	ChatGPT Label	Word description
14	3.71%	344973	Chemotherapy and Treatment Outcomes	chemotherapy, toxicity, cisplatin, chemotherapy, metastasis, gemcitabine, progression_free_survival, adverse_event, docetaxel, paclitaxel, bevacizumab, treatment, fu, overall_survival, platinum, phase_iii
15	3.61%	329108	Genetic Mutations and Epigenetic Changes	mutation, DNA_methylation, genome, chromosome, sequence, deletion, allele, colorectal_cancer, exon, iras, mutations, PCR, epigenetic, amplification, molecule
16	3.60%	308435	Blood Cancer	transplant, acute_myelogenous_leukemia, leukemia, leukemia, bone_marrow, hematopoietic_stem_graft, vs, host_disease, chronic_lymphocytic_leukemia, donor, stem, relapse, myeloma, chd1, brain_metastases, acute_lymphoblastic, chronic_myeloid_leukemia
17	3.54%	302595	Medical Imaging	imaging, computed_tomography, nuclear_magnetic_resonance, positron_emission_tomography, scan, biopsy, benign, nodules, radiology, emission, ultrasound, predictive_value, perfusion, preoperative, staging
18	3.16%	329746	Neurological Aspects and Brain Tumors	cerebral, glioma, neuron, central_nervous_system, glioblastoma_multiforme, glioblastoma, rat, cardiovascular, cerebrospinal_fluid, alzheimer_disease, virus, dexamethasone, seizure, meningitis, heart, temozolomide
19	3.09%	315020	Viral Infections and Cancer	human_papillomavirus, melanoma, cutaneous, cervical_lymphoma, squamous_cell_carcinoma, gastric, Epstein_Barr_virus, non_hodgkin_lymphoma, helicobacter_pylori, pylori, esophageal_biopsy, mucosa, cytology
20	3.06%	326583	Inflammation and Polymorphisms	polymorphism, tumor_necrosis_factor, serum, inflammation, plasma, genotype, rheumatoid_arthritis, necrosis_factor, renal_cytokine, allele, protein, blood, C_reactive_protein, vitamin_d
21	2.68%	398123	Specific Cancer Types and Therapies	breast, lung, epidermal_growth_factor, non_small_cell_lung_cancer, estrogen_receptor, triple_negative_breast_cancer, parkinson_disease, metastasis, tamoxifen, estrogen, circulating_tumor_cells, receptor_tyrosine_kinase, adenocarcinoma, melanoma, trastuzumab
22	2.66%	285492	Endocrine and Hormonal Factors	ovarian_cancer, thyroid_endometrial_cancer, insulin, papillary_thyroid_carcinoma, polycystic_ovary_syndrome, pregnancy, hormone, pituitary, obesity, meformin, uterine, polycystic_ovary, of, follicular
23	2.59%	233468	Radiotherapy and Treatment Techniques	radiationtherapy, radiation, Gray_Intensity_Modulated_Radiation_Therapy, imaging, docetaxel, beam, phantom, computed_tomography, brachytherapy, toxicity, tissue, Stereotactic_Body_Radiation_Therapy, energy, proton
24	1.70%	202626	Prostate Cancer and Treatment Approaches	prostate, prostate_cancer, bone, prostate_specific_antigen, androgen_receptor, androgen, post_radical_prostatectomy, biopsy, gleason_score, metastasis, antibiotic, Androgen_Deprivation_Therapy, hcup, cells, androgen_deprivation, osteoclast
25	1.69%	209151	Liver Cancer	hepatocellular_carcinoma, liver, hepatitis_b, hepatitis_c, hepatocyte, alpha_fetoprotein, cholangiocarcinoma, hb_surf_cancer, omissis, radiotherapy, ablation, TACE, chronic_hepatitis, hepatosty, sorafenib, gaba/gabader

