

Projet « Patent2Net »

Le « pivot CIB » - bases académiques

Reymond, D. IMSIC (Institut Méditerranéen des Sciences de l'Information et de la Communication) EA 7492, Université de Toulon, Aix Marseille Univ, Toulon, France

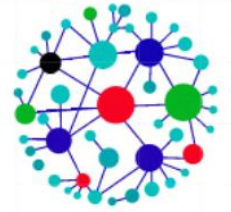
Patent2Net

- Depuis 2014
- Utilisé au Brésil, Mexique, Allemagne, Suisse, Sénégal
- Collecteur de données brevet (texte, image)
- **Générateur de corpus**
- Terrains de pratique d'outils de DM et de formation aux humanités numériques (Reymond & Noyer, 2022)

Sur deux projets 2019-2022 à Financement GIS-URFIST (11k€)

1. Stagiaire (version docker)
2. Stagiaire (avec LEAD) qualification d'auctorialité

Patent2Net

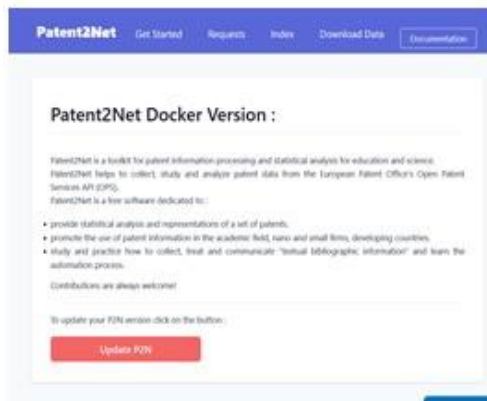


Study and practice how to process,
analyze and communicate on data
using patents and data-mining



GIS "RÉSEAU URFIST"

Groupement d'intérêt scientifique



localhost:5000

Flask
app.py

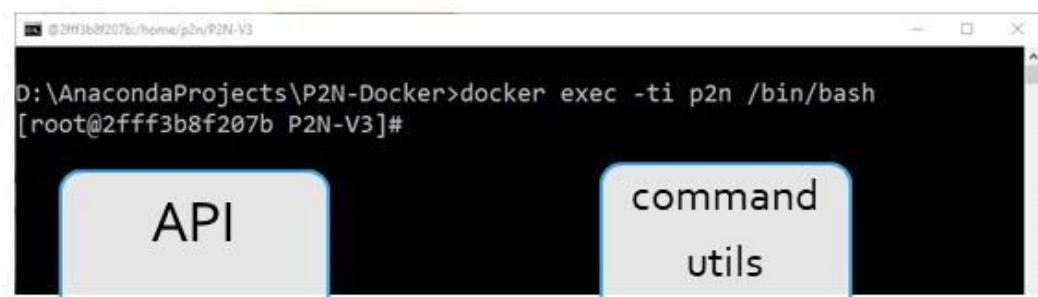
Main scripts
Patent2Net

Request splitter helper

Gathering

Pre-processing

Processing



API
p2n

command
utils
models...

Libraries

Shared functions

Templates
Web interface
formatting

Static
information

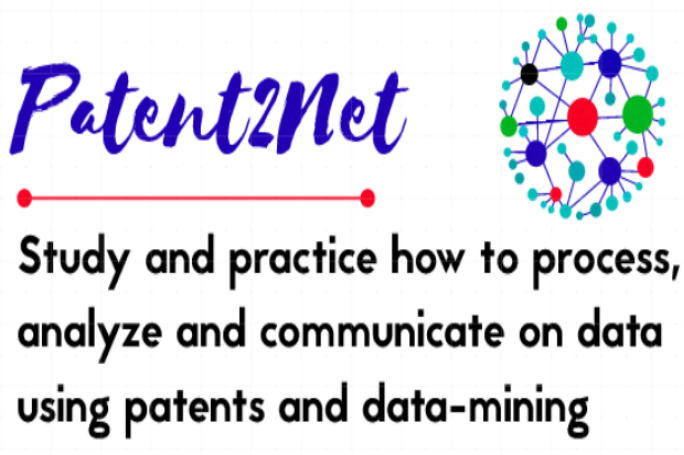
SymbolesIPC.csv
International Patent
Classification symbols

NamedCountryMap.csv
Country information

lpcr_2015.xml. IPC
symbols description

StanNorm2.csv (adapted from EPO).
Applicant's names normalization dictionary

Welcome to Patent2Net



Python 3.6 Github stars 9 issues 3 open docs passing

About

Hint
 Patent2Net is a tool suite for patent information processing and statistical analysis for education and science. Patent2Net aims to collect and help study and analyze patent data from the European Patent Office's [Open Patent Services API \(OPS\)](#).

Patent2Net is free software, dedicated to:

- provide statistical analysis and representations of a set of patents.
- promote the use of patent information in academic, nano and small firms or developing countries
- learn, study and practice how to collect, process and communicate "textual bibliographic information", and automation process
- learn several information processing skills in the *Library and Information Science* fields.
- learn skills in data-mining software, Data analysis, Textual data-mining, distance reading, knowledge discovery
- See extra-feature with [the docker version](#)

passiflora

ta=passiflor* or ta="passion fruit**

08 Jun 2022

Number of patents retrieved	2826	Number of family patents retrieved	3155
Abstract: OL: 2139 EN: 2812 DE: 11 FR: 34		FamiliesAbstract: OL: 2063 EN: 2822 DE: 17 FR: 91	
Claims: DE: 8 PT: 5 ES: 1 RO: 1 FR: 6 ZH: 3 EN: 19		FamiliesClaims: DE: 28 JA: 3 NO: 1 PT: 9 HR: 2 ES: 14 RO: 1 FR: 35 ZH: 4 EN: 94 IT: 1	
Description: DE: 8 PT: 5 ES: 1 RO: 1 FR: 6 ZH: 3 EN: 19		FamiliesDescription: DE: 15 JA: 3 NO: 1 PT: 9 ES: 15 RO: 1 FR: 23 ZH: 4 EN: 90 IT: 1	

On-line analysis tools

Patents

Patents datatable [↗](#)
Explore patents in table format

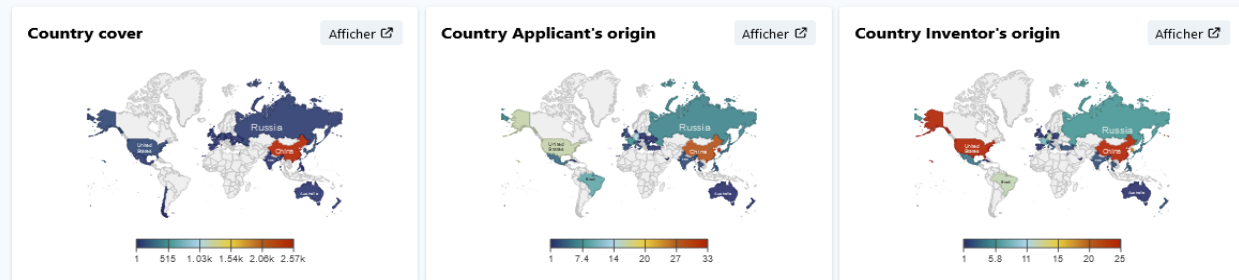
Pivot table [↗](#)
Build your own indicators crossing several fields

Images disponibles [↗](#)

Networks analyser tool [↗](#)
Explore and draw co-(inventors, applicants or technology) networks

Carrot ES Plugin [↗](#)
Cluster documents with the awesome carrot2 tool

Attractivity



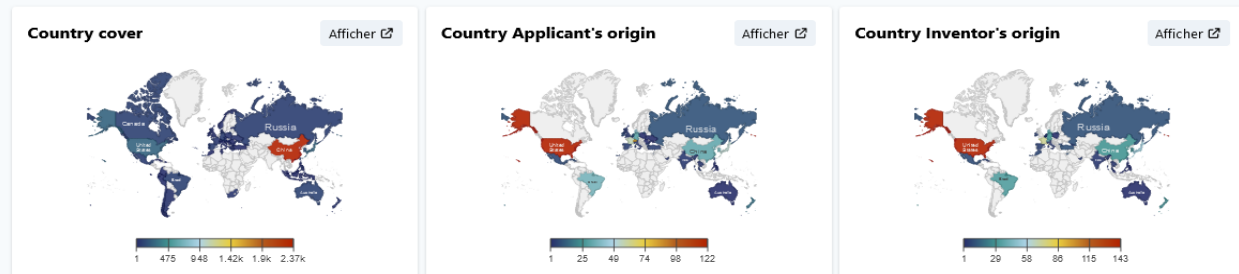
Familles

Patents datatable [↗](#)
Explore families in tables (fix grouping patent in families)

Pivot table [↗](#)
Pivot indicators for families

Images from patents [↗](#)

Attractivity



Patent2Net
latest

SETUP

Quick start

USAGE

Usage

AVANCED SETUP

Historical features

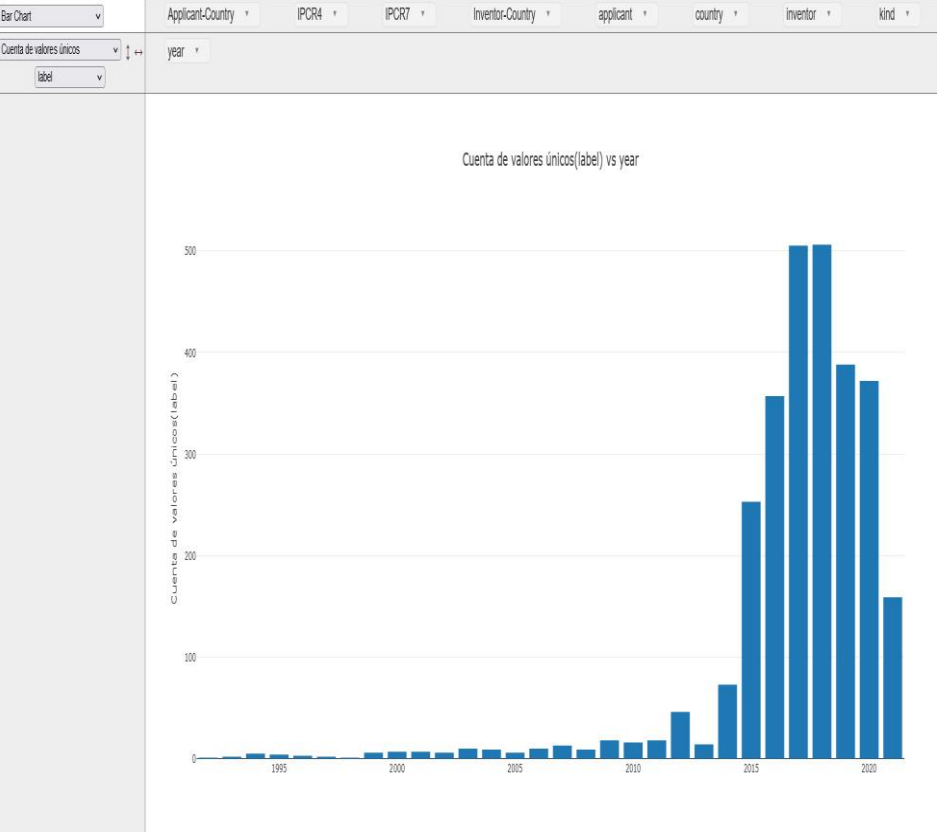
CORE INFORMATION

Global architecture and data model

DEVELOPMENT INFORMATION

Development info

Indicateurs et cartographies



Heatmap

Applicant-Country | IPCR7 | Inventor-Country | applicant | inventor | kind | label | prior-Date

Cuenta de valores únicos (label) vs country

country	IPCR4	A01B	A01C	A01D	A01F	A01G	A01H	A01K	A01M	A01N	A01P	A21C	A21D	A23B	A23C	A23D	A23F	A23G	A23K	A23L	A23N	A23P	A24B	A24D	A24F	A47C	A47F	A47J	A61J	A61K	A61L	A61M	A61P	A61Q	B01D	B01F	B01J	B01J					
AU																				2										1		1											
BE																														1													
BG																														3			1										
CL																																							1				
CN	18	12	44	11	2	302	39	10	15	22	29	2	61	43	77	5	136	84	78	60	93	24	7	2		1	1	16	1	178	4	1	294	75	14	5	1						
CZ																																											
DE																					2										5					2	4						
EA																		1																									
EP																				2										5			2			1							
ES																														1													
FR							1																															1					
GB																																											
GE																																											
IE																																											
IL	1																																										
IN	1																																										
JP																			2			16									10			34	24					1			
KR											2										9									16			8	10									
MX											1								1	1	1									3			2										
NL																																											
NZ																						1									3						1						
PH																																											
PL																						1																					
RO																																											
RU																																											
TW							1						1								4	1								6											2		
UA																																											
US								2																																			
WO																																											
Totales	20	12	44	11	2	304	41	10	15	25	29	2	62	44	77	6	136	91	79	60	94	24	8	2	2	1	1	16	1	278	4	1	357	129	21	5	4						

Réseaux et lecture distante

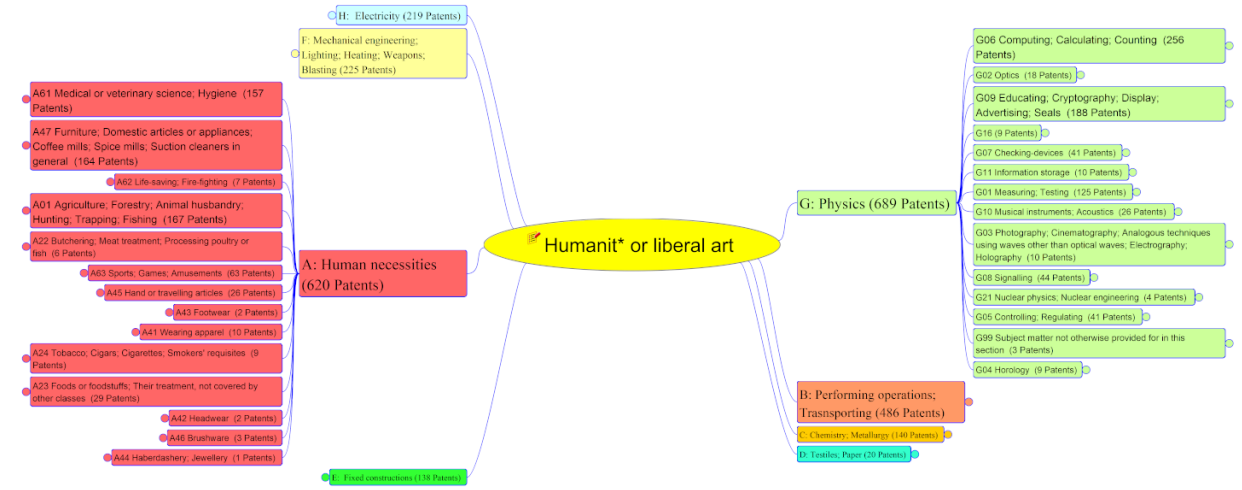
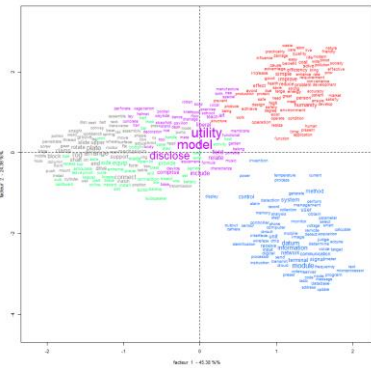
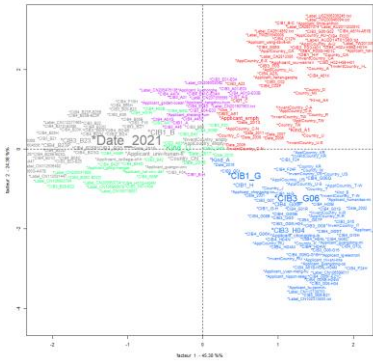
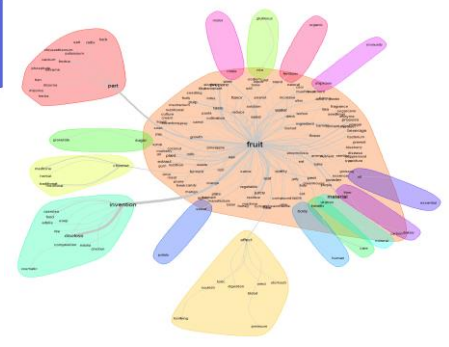
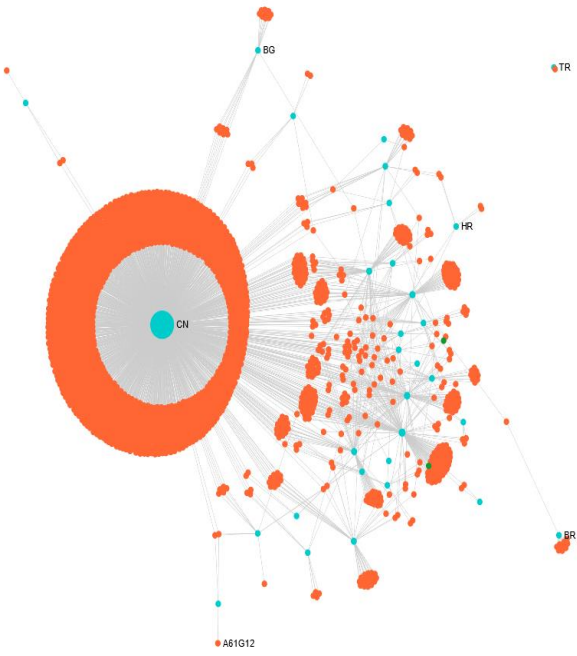
Draw graph
Dynamic visualization of your graph

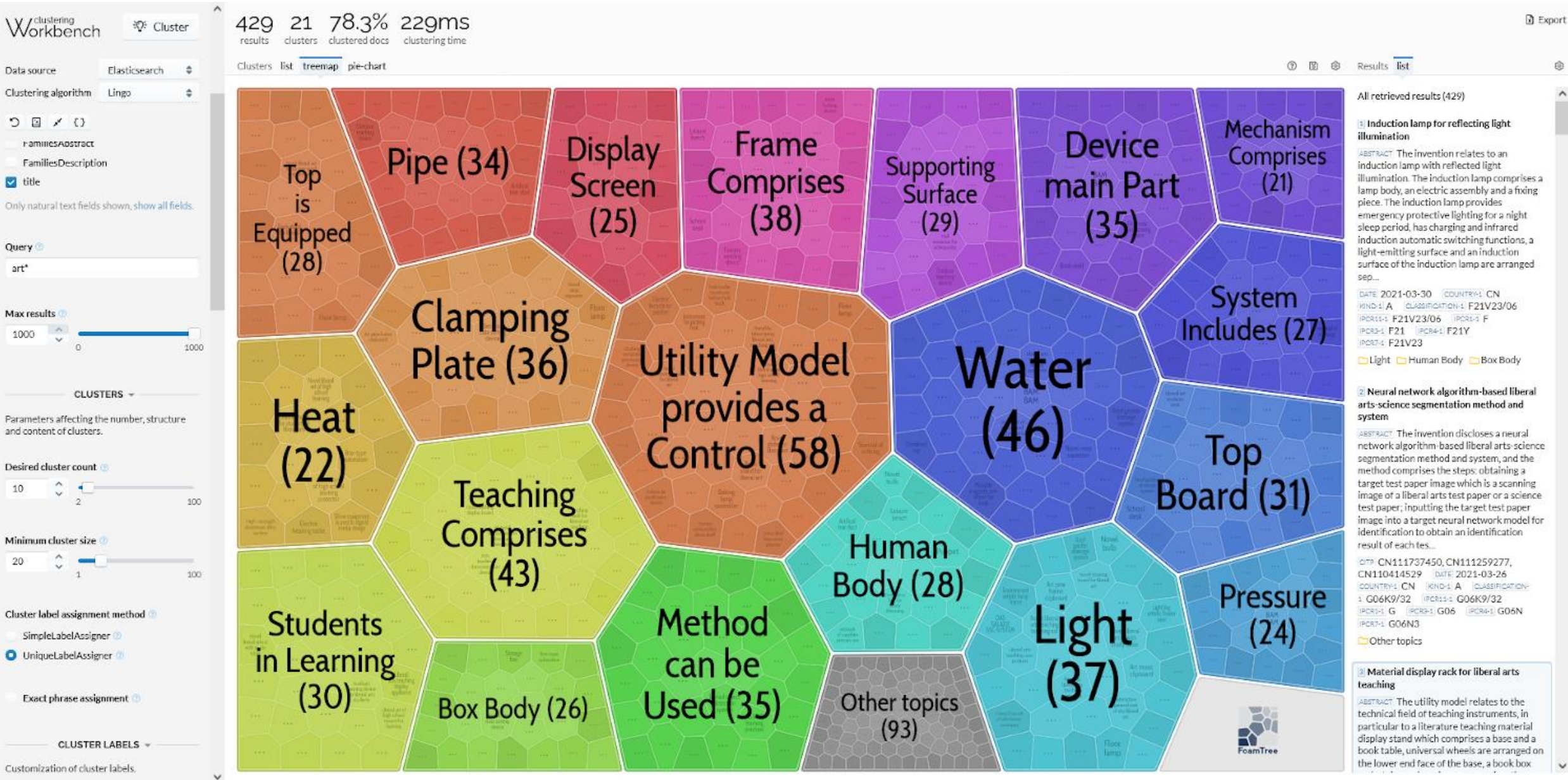
← Return to recipe ×

Start Layout/Rescale

Legend
Color by
category

- country
- IPCR7
- IPCR4





Potentiel

- Intégration de collecteurs de production académique
- ElasticSearch(Kibana) en alternative aux différents outils de cartographie
- Réalisation de supports video pour formation

Les collaborations science industrie

Retrouver l'inventeur-auteur et la levée d'homonymies d'autorat
entre les brevets et les publications scientifiques

Reymond, D. (IMSIC), Khouilla, H. (LEAD), Durand-Barthez, M.
(URFIST Paris), & Wolff S. (BETA)

La relation science indu

- Les brevets co-déposés par un organisme de recherche public et un organisme privé
 - Identifier des motifs de ces collaborations
 - Tracer des spécificités
 - [...]
- Corpus de brevets
- Mais il faut :
 - la liste des établissements publics
 - être sûr que ce co-dépôt soit trace de toutes les interactions
 - I.e. est-ce que des inventeurs, au titre d'un organisme privé, ne seraient pas issus de la recherche publique ?

CNRS (CENTRE NATIONAL DE LA RECH SCIENTIFIQUE)
CENTRE NAT DE LA RECH SCIENTIFIQUE(CNRS)
CT NAT DE LA RECH SCIENTIFIQUE
CNRS CENTRE NATIONAL DE LA RECH SCIENTIFIQUE
CENTRE NAT DE LA RECH SCIENTIFIQUE CNRS
CNRS DIRE
CT NAT DE LE RECH SCIENT
SCIENT C N R S CT NAT DE LA RECH
CT NAT DE LA RECH CNRS
LE CENTRE NAT DE LA RECH SCIENT
CT NAT DE LA RECH SCIENTIFIQUE—CNRS—
SANTR NASONAL DE LJA RESHERSH SJANTIFIK
CT NAT DE LA RECH SCIENTIFIQUE—CNRS
CENTER NAT DE LA RECH SCIENT C N R S
CT NAT DE LA RECH SCIENT PUBLI
PHARMAMENS E CT NAT DE LA RECH SCIENT CNRS
ENTRE NAT RECH SCIENT
CT NAT DE LA RESERCHE SCIENT CNRS
CENTRE NATIONAL DE LA RECH SCIENTIFIQUE—CNRS
CENTRE NAT DE LA RECHERCH SCIENT
EHNSTITJU NAS ONAL DE LJA RESH
CENTRE NATIONAL DE LA RECH SCIENTIFIQUE—CNRS

Le problème : lien science industrie ou pas...

☆ AU2015208975A1 TiO2 material that is absorbent in the visible range and method for producing same

Disponible en ▾ Patent Translate ▾ ⋮ ✕

Données bibliographiques ▾

Dossier mondial ↗

Demandeurs ENSCM; CENTRE NAT RECH SCIENT; TOTAL SA - ECOLE NATIONALE SUPERIEURE DE CHIMIE DE MONTPELLIER; CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE - CNRS; TOTAL SA

Inventeurs BOSCARO PAOLO; HULEA VASILE; MARCOTTE NATHALIE; FAJULA FRANCOIS; GALARNEAU ANNE; LUCK FRANCIS +

Classifications

CIB **A61L9/20; B01J21/06; B01J35/00; B01J35/02; B01J35/10; B01J37/03; C01B3/04; C02F1/32;**

☆ CN1394594A Amino acid derivative base compound and application and method for dissolving material with low solubility

Disponible en ▾ ⋮ ✕

Données bibliographiques ▾

Demandeurs OREAL [FR] +

Inventeurs GIRAMAN CAROLE [FR]; CARPURAN DOMINIK [FR]; SCHELYWAILLE WIKRONIK [FR] +

Classifications

CIB **A61K31/506; A61K31/5685; A61K31/616; A61K8/00; A61K8/04; A61K8/14; A61K8/30; A61K8/33; A61K8/37; A61K8/40; A61K8/44; A61K8/49; A61K8/63; A61P17/00; A61P17/10; A61Q1/00; A61Q1/02; A61Q19/00; A61Q19/02; A61Q19/08; A61Q19/10; A61Q5/00; A61Q5/08; (IPC1-7): A61K7/06; A61K7/40; A61K7/48;**

Hypothèse

→ **L'inventeur d'un brevet qui publie est un inventeur-auteur potentiel**

- Donc à un (nom, prénom) si l'on retrouve des publications dans une base documentaire du domaine, on a un soupçon

Mais : reste à identifier s'il ne s'agit pas d'un homonyme **et si ce n'est pas le cas de** retrouver son affiliation

Contraintes

→ Demander la carte d'identité est un problème difficile (voire insoluble),

→ Une base de tous les chercheurs du monde pas encore disponible (ne résoudrait pas l'homonymie directement même si cette base inclus la liste des publications dudit chercheur)...

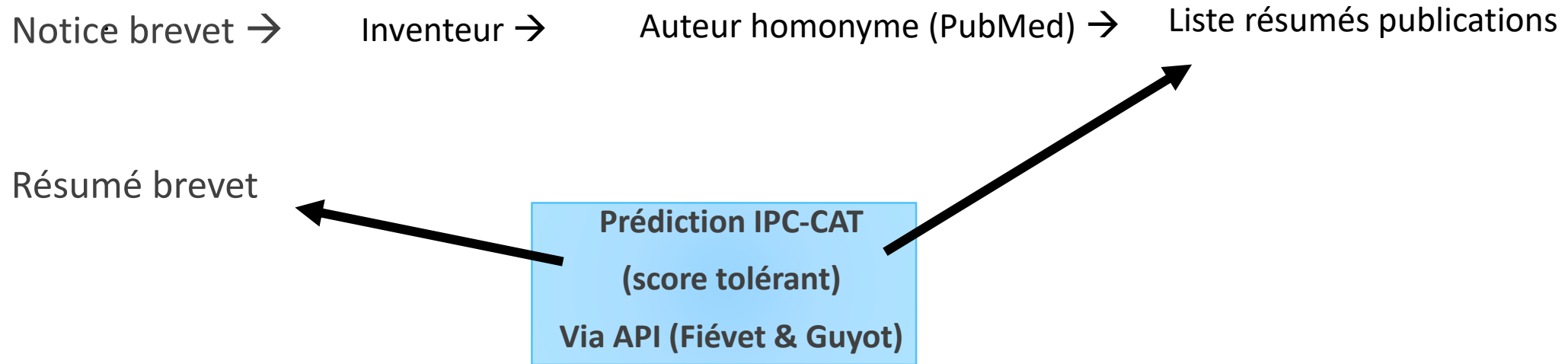
→ Une ou des publications dans un domaine proche du brevet renforce le soupçon

Méthodologie

3 corpus :

- 1 « pur académique » : contrainte sur les déposants (Univ*), priorité FR (683 demandes brevets)
- 1 global sur le même domaine technologique (3122 brevets)
- 1 sur un autre domaine technologique (874 brevets)

Procédure



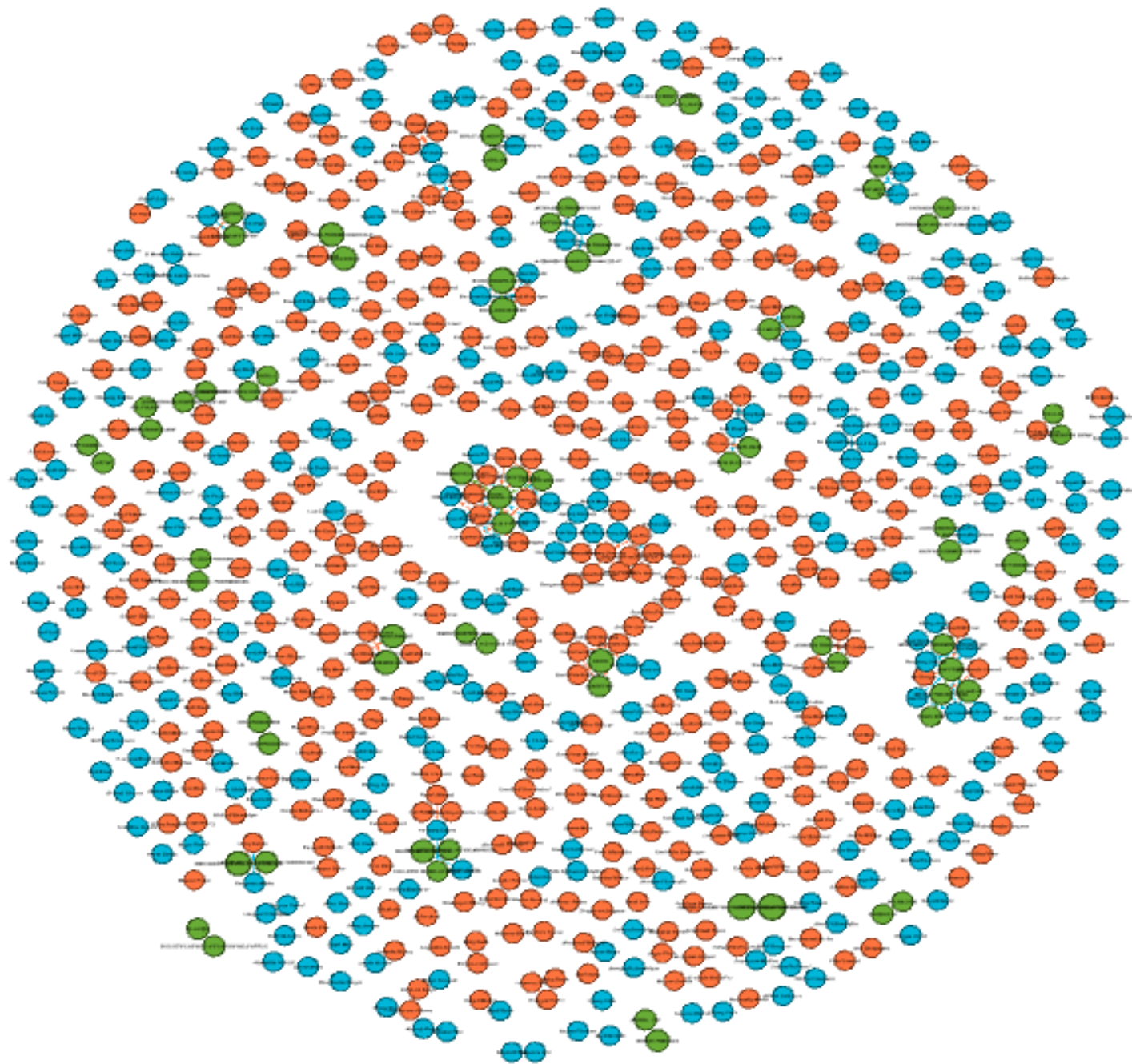
Si au moins un MATCH alors l'homonyme n'en est probablement pas un

Résultats

corpus	Univ	Large	LargeBis	Total
Brevets collab (>1 déposant)	672	596	2494	3762
Inventeurs	2122	1088	4510	7720
Auteurs candidats (match IPC > 800)	1209	141	1265	2501
Echantillon aléatoire	137	14	244	390
Match	133 (97%)	14	231 (95%)	373 (96%)
Doute	3	0	12	15
Erreur	1	0	1	2 (0.005%)

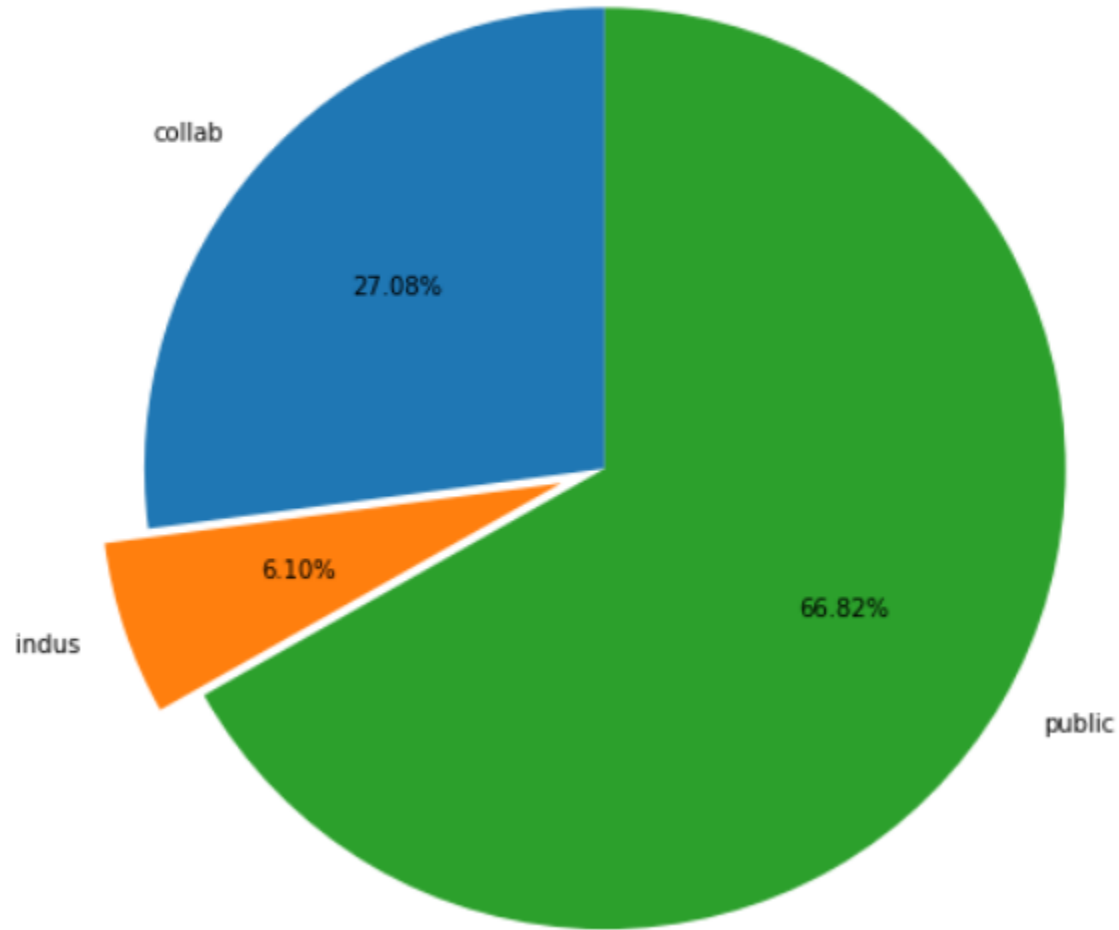
Réseau de co-dépôts
Auteurs **affiliés univ**
fr, **non affiliés fr** et
sociétés privées

soit environ 20% du
sous-corpus
industriel LargeBis :
≈1000 auteurs
71 sociétés

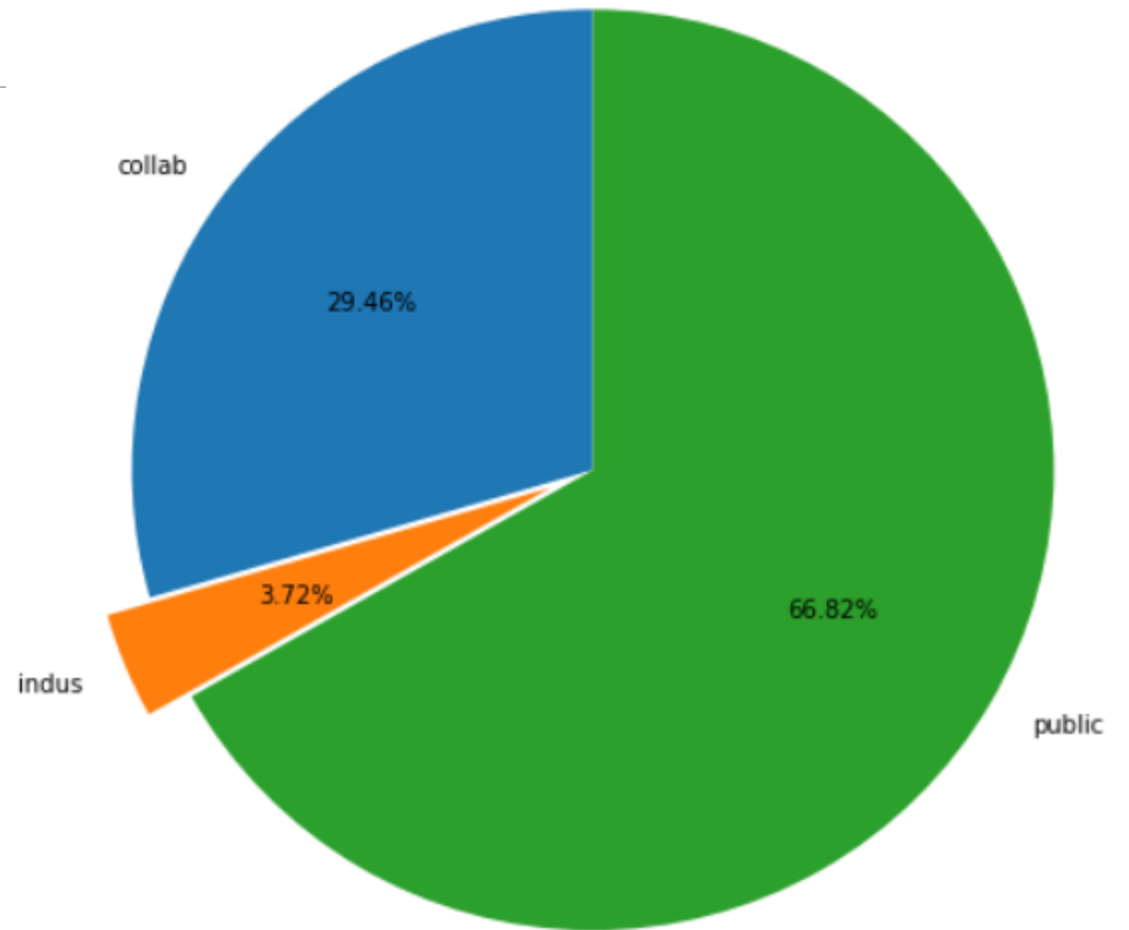




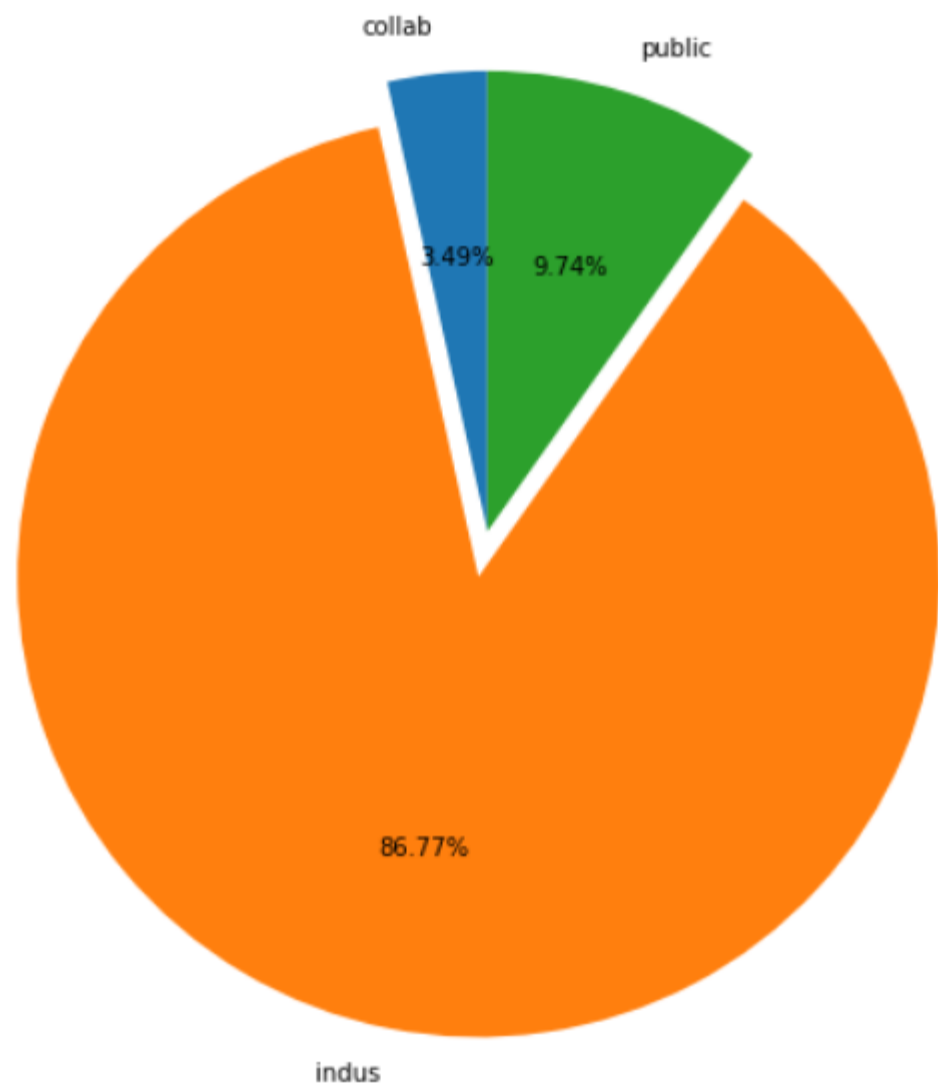
Univ (672 brevets)



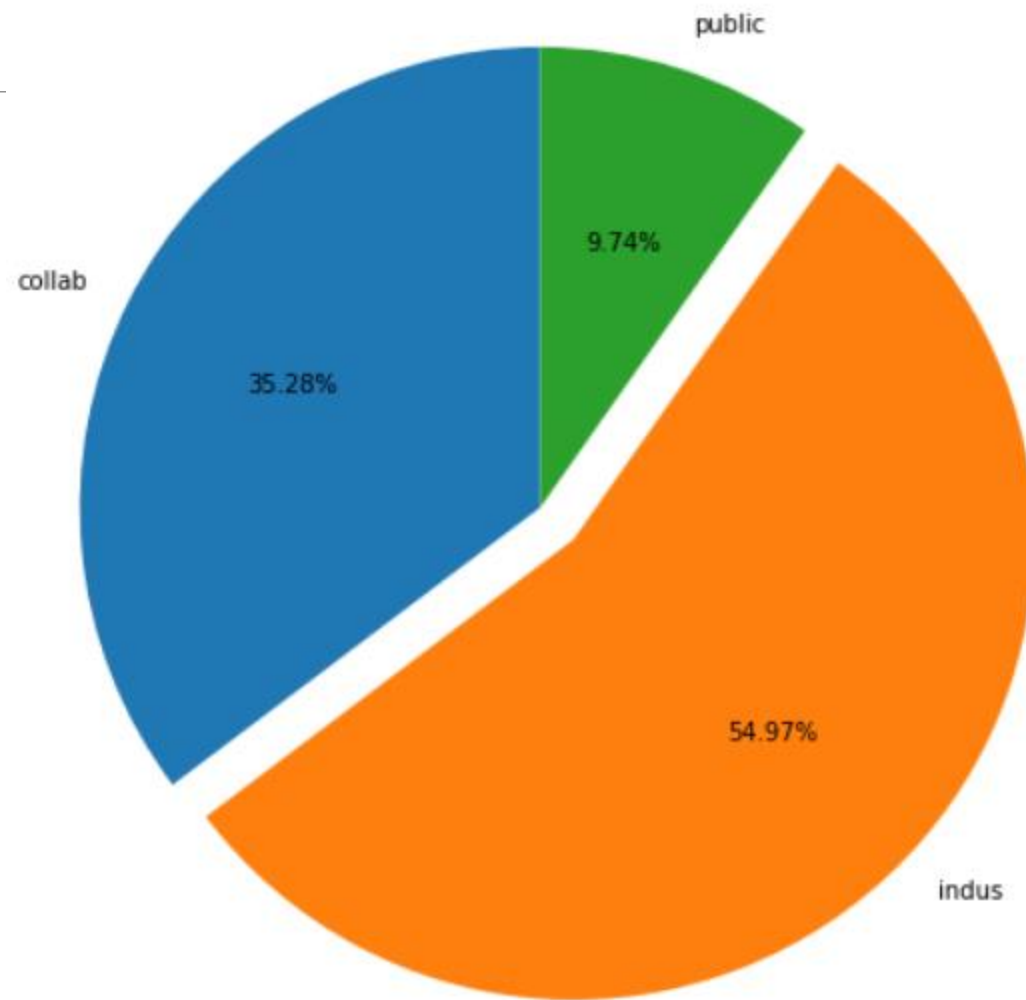
Répartition collaborations incluant affiliations



Largebis (2494 brevets)



Répartition collaborations incluant affiliations



Conclusion

Cf. Reymond (et al., 2022)

- Affine les indicateurs « traits de la collaboration science-industrie »
- Extensible à d'autres domaines technologiques (si BDD dispo)

Travaux futurs

- Analyser les données de collaboration affinées
- Pose la question de la suffisance des brevets pour la reconstitution d'arbres technologiques : quid des développements académiques qui pourraient être des briques sous-jacentes

Références

Reymond, D., Khouilla, H., Durand-Barthez, M., & Wolff, S. (2022, paraître). *Retrouver l'inventeur-auteur : La levée d'homonymies d'autorat entre les brevets et les publications scientifiques*. JADT2022: 16th International Conference on Statistical Analysis of Textual Data, Naples, Italy.

Reymond, D., & Noyer, J.-M. (2022). Se former aux « Humanités numériques » par l'Open Data : Exemples par les chaines infométriques de Patent2Net (P2N). *Humanités numériques*, N°5. <https://doi.org/10.4000/revuehn.3014>

Fiévet, P., & Guyot, F. (2018). *Automatic Categorization of Patent Documents in the International Patent Classification (IPCCAT)*. The International Conference on Search, Data and Text Mining and Visualization. (IC-SDV), Nice. <https://haxel.com/ii-sdv/2018/Programme/monday-23-april-2018>