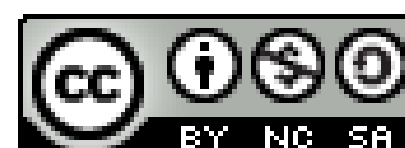
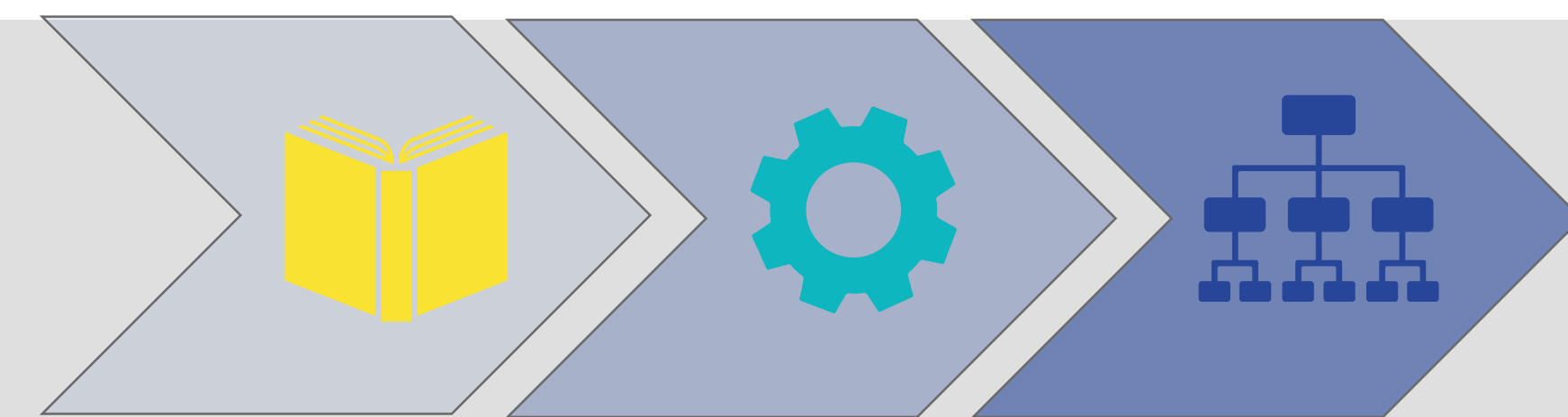
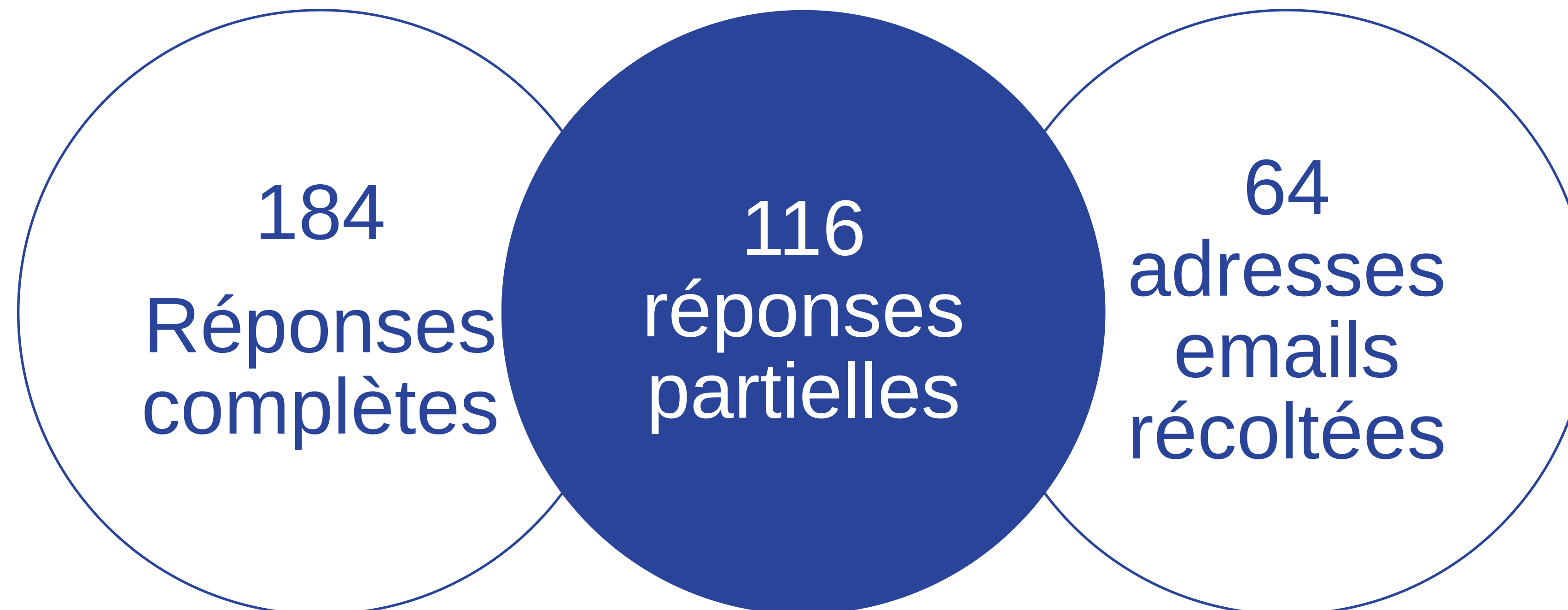


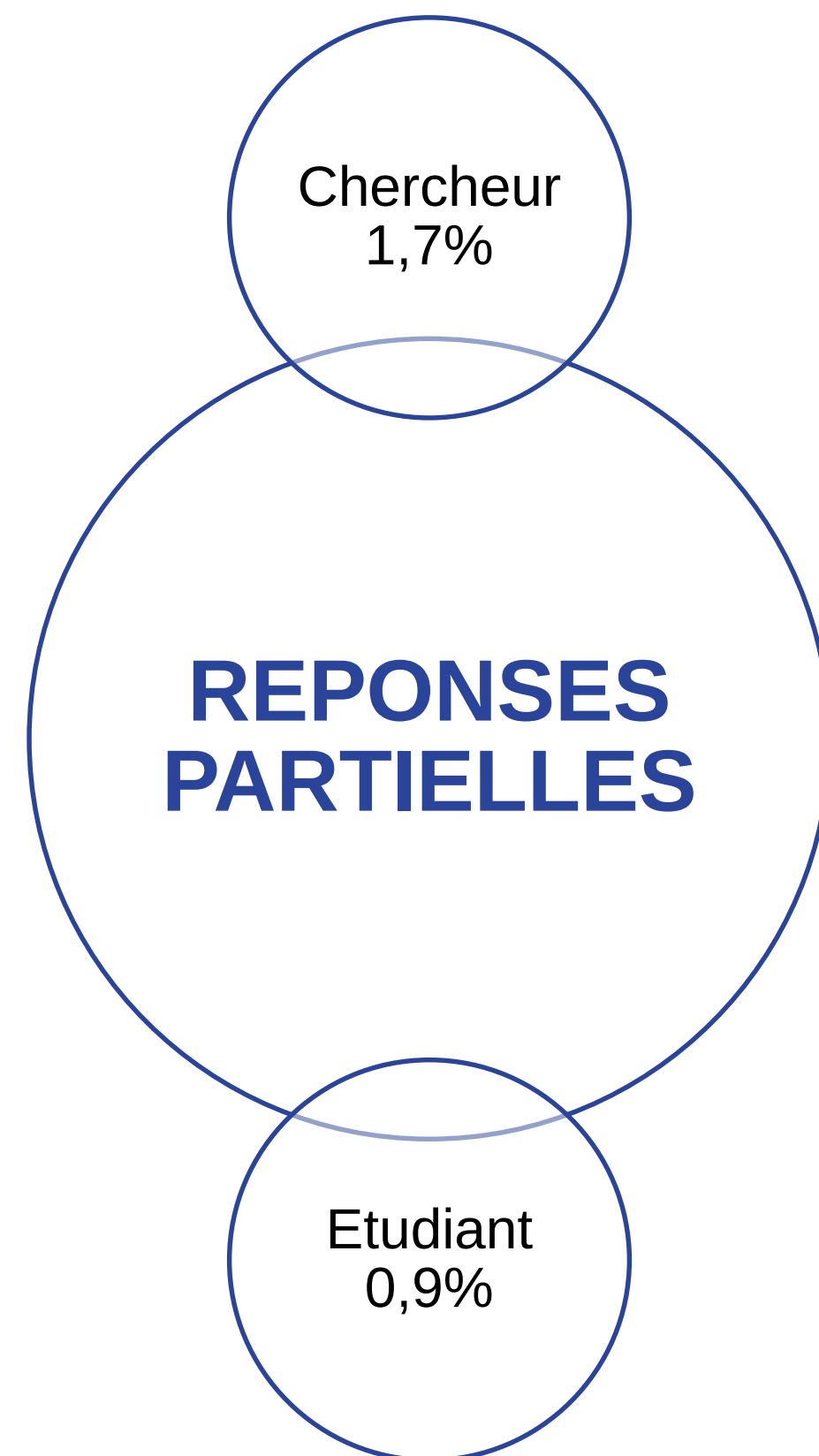


VisaTM étude

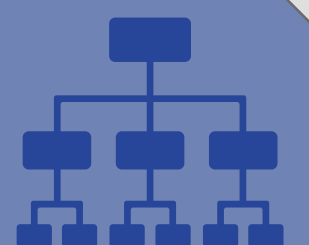
Résultats de l'enquête Visa TM





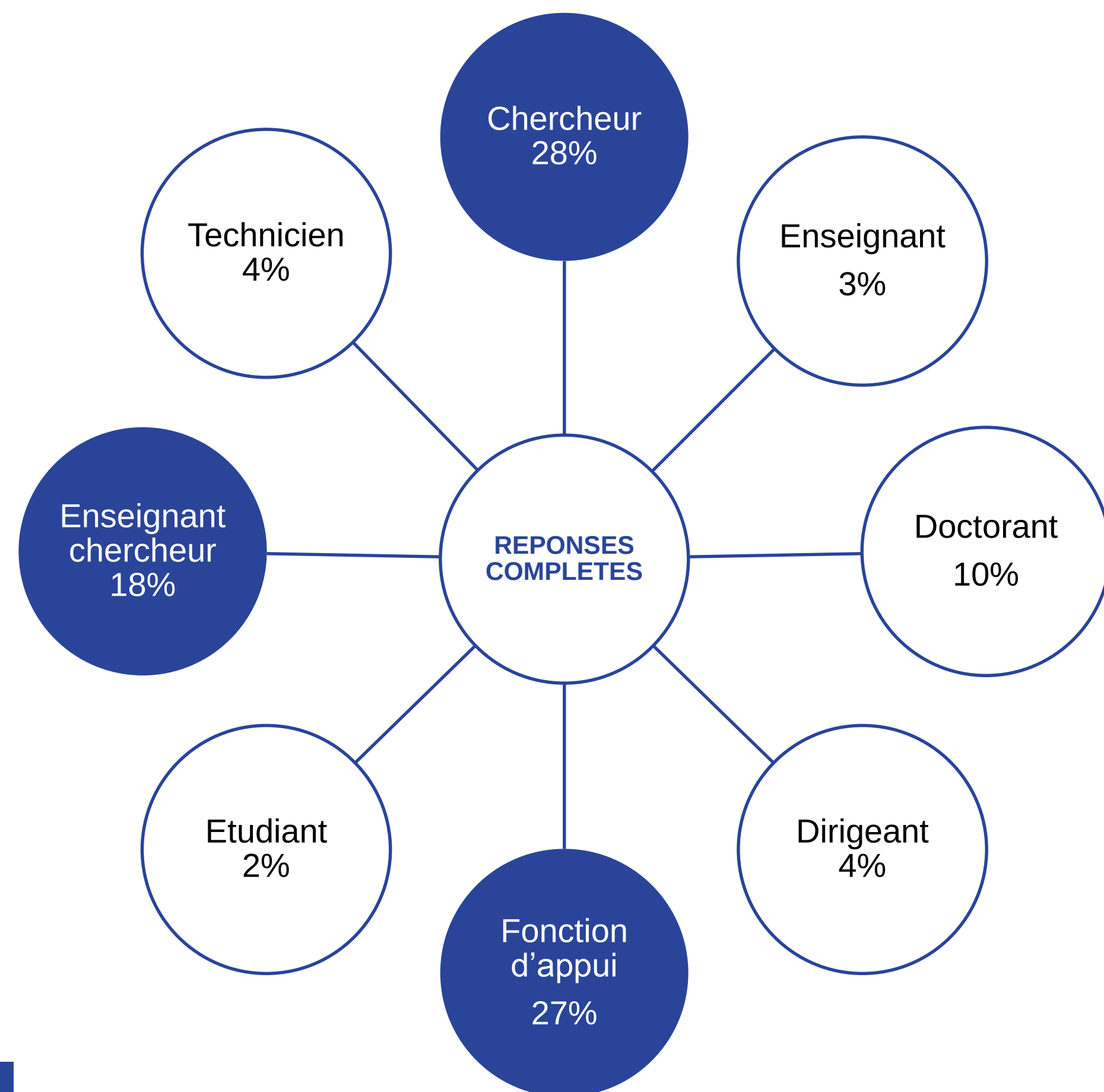


116
Réponses partielles



Réponses complètes / typologie des utilisateurs

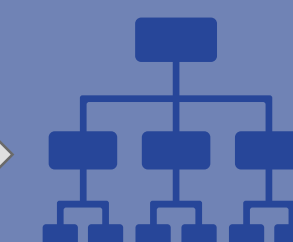
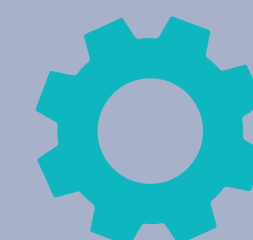
300
réponses



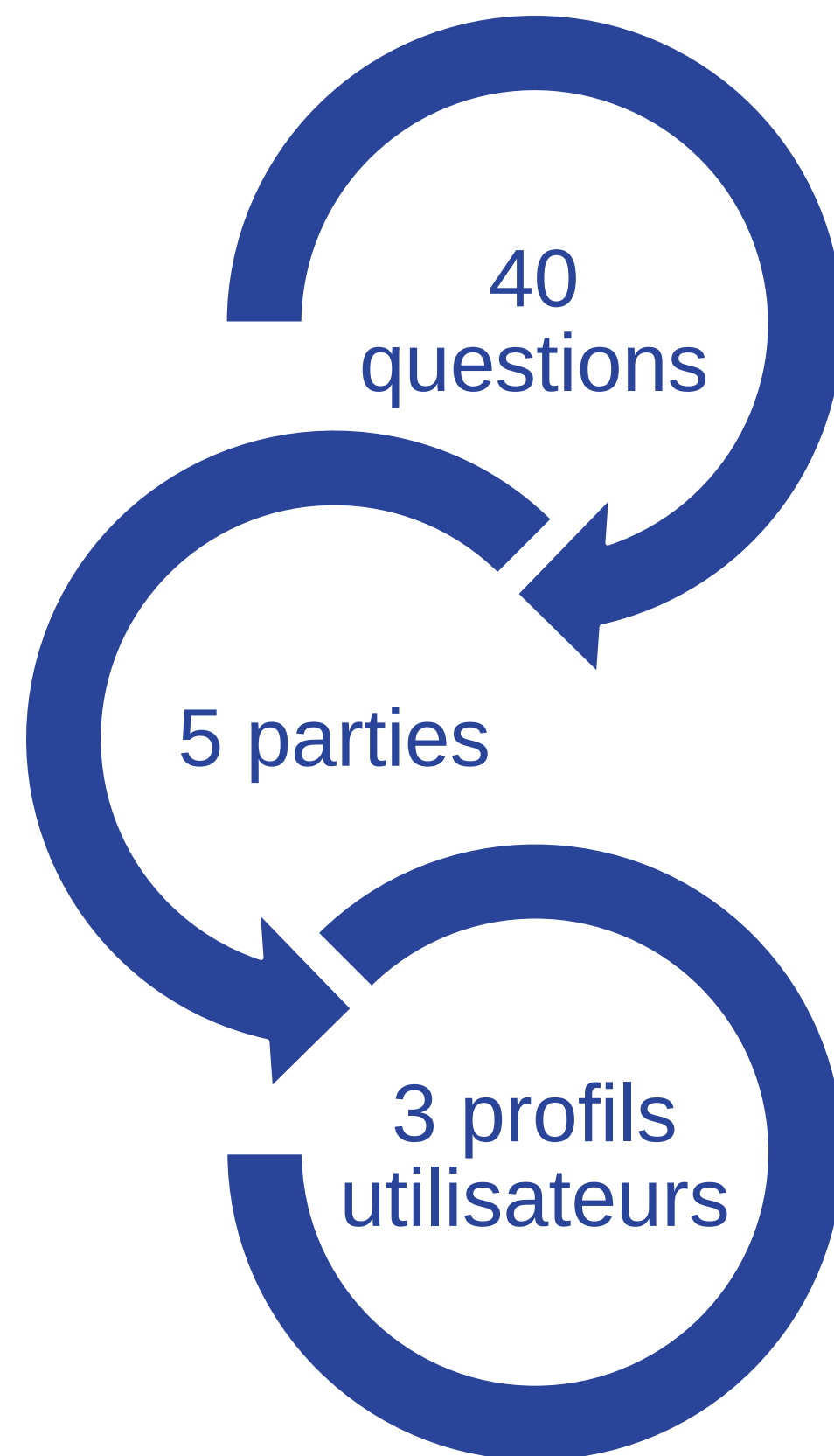
184
Réponses complètes

AUTRE

Chef de projet, ingénieur d'étude, ingénieur de recherche, ingénieur data, documentaliste, consultant département R&D, chargé de projet dans les outils numériques, architecte, développeur



Structure du questionnaire



Partie 1

La fouille de textes dans vos activités

Partie 2

Compétences et formation

Partie 3

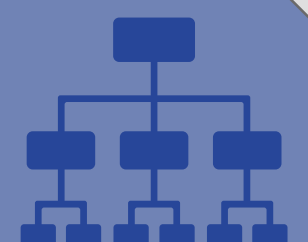
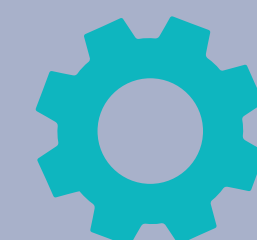
Freins et leviers

Partie 4

Comment verriez-vous les fonctionnalités d'une fouille de textes

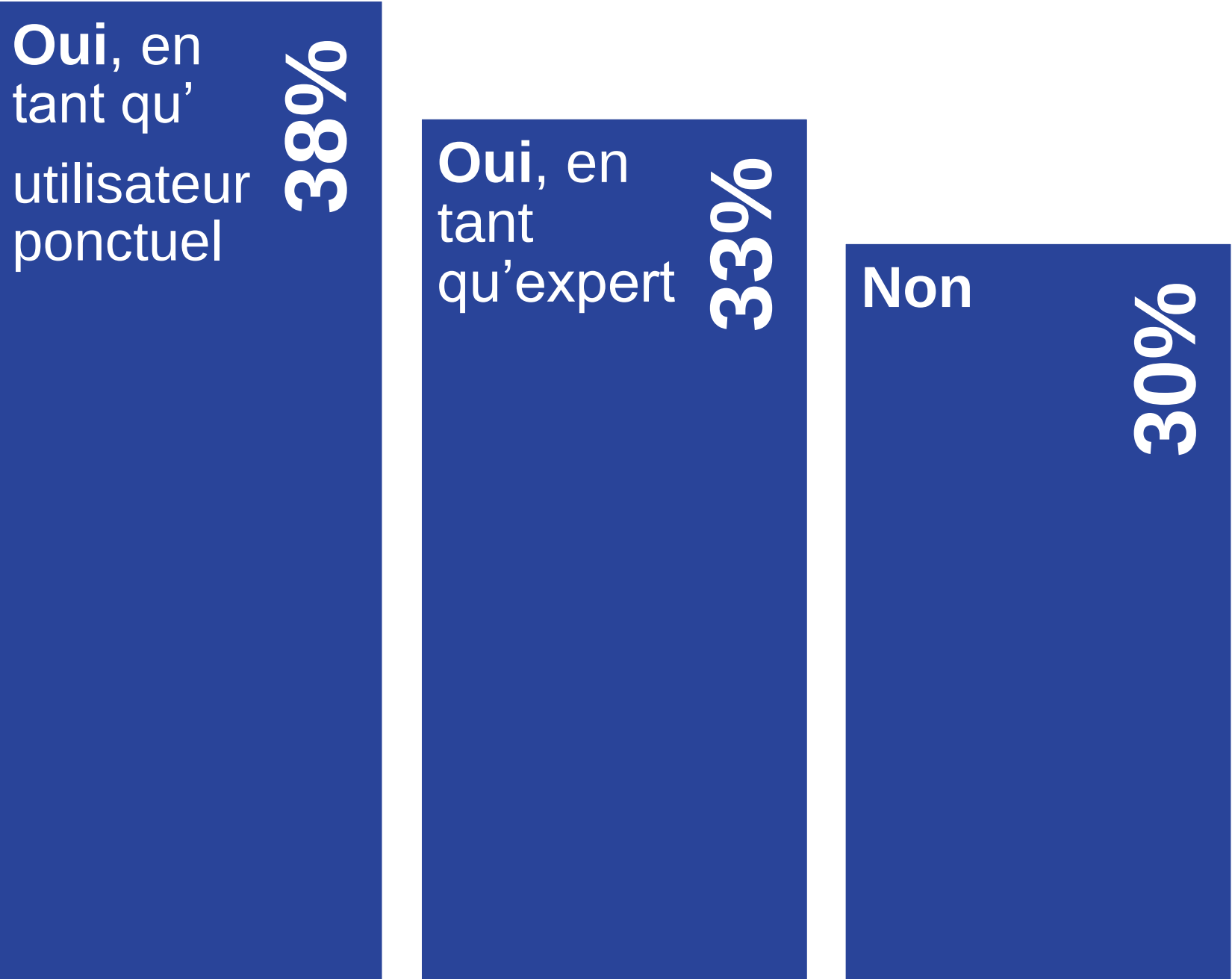
Partie 5

Vous



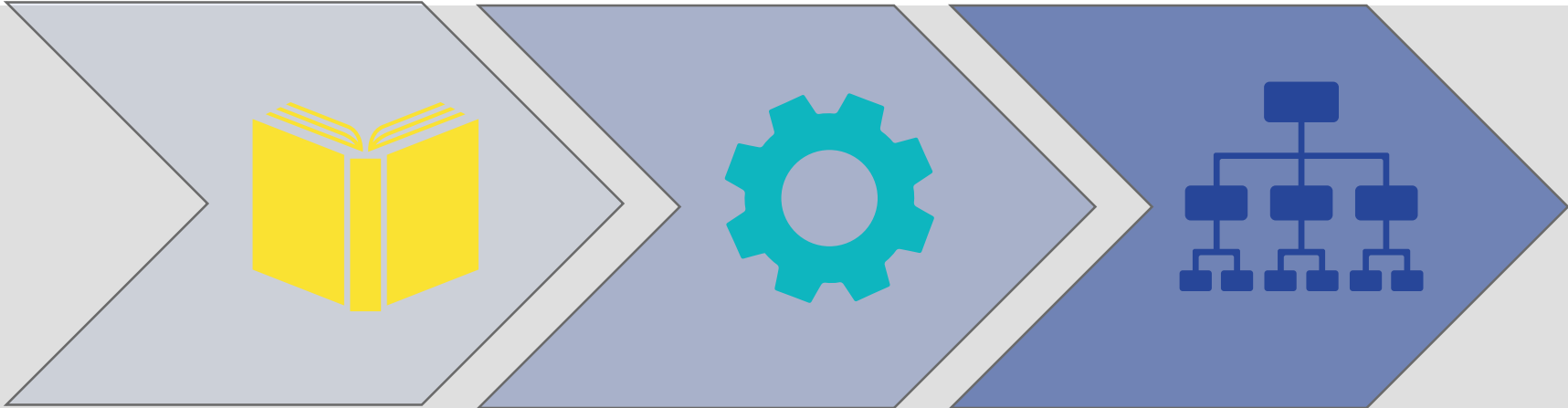
La fouille de texte dans vos activités

Selon cette définition, pouvez-vous dire que vous pratiquez ou faites appel à la fouille de texte ?

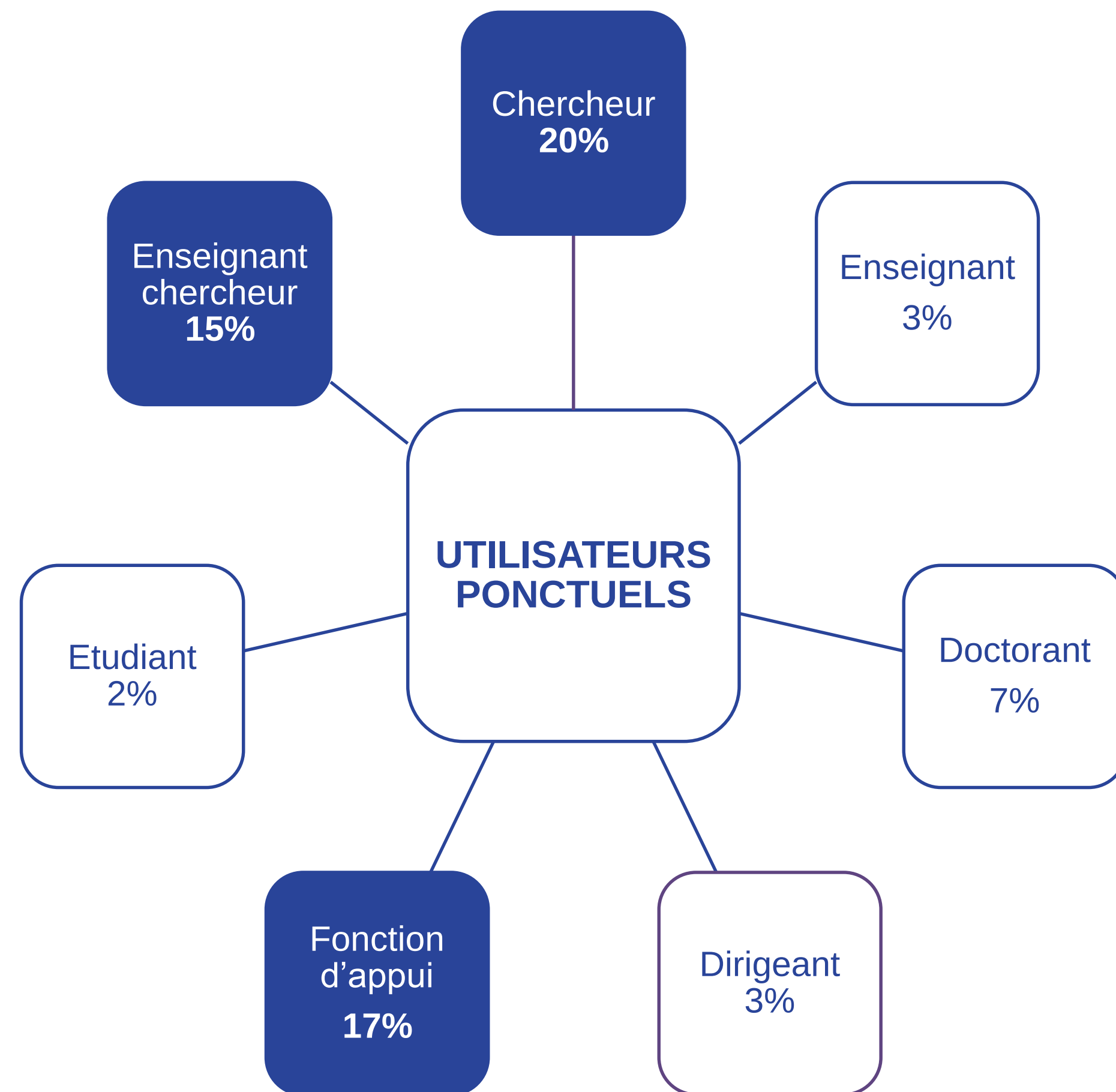


- 38%
Oui, en tant qu'utilisateur ponctuel
- 33%
Oui, en tant qu'expert
- 30%
Non

273 réponses



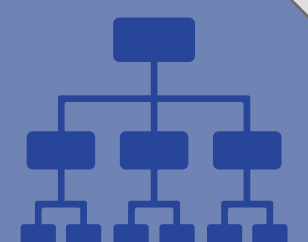
Pratique de la fouille de texte / utilisateur ponctuel



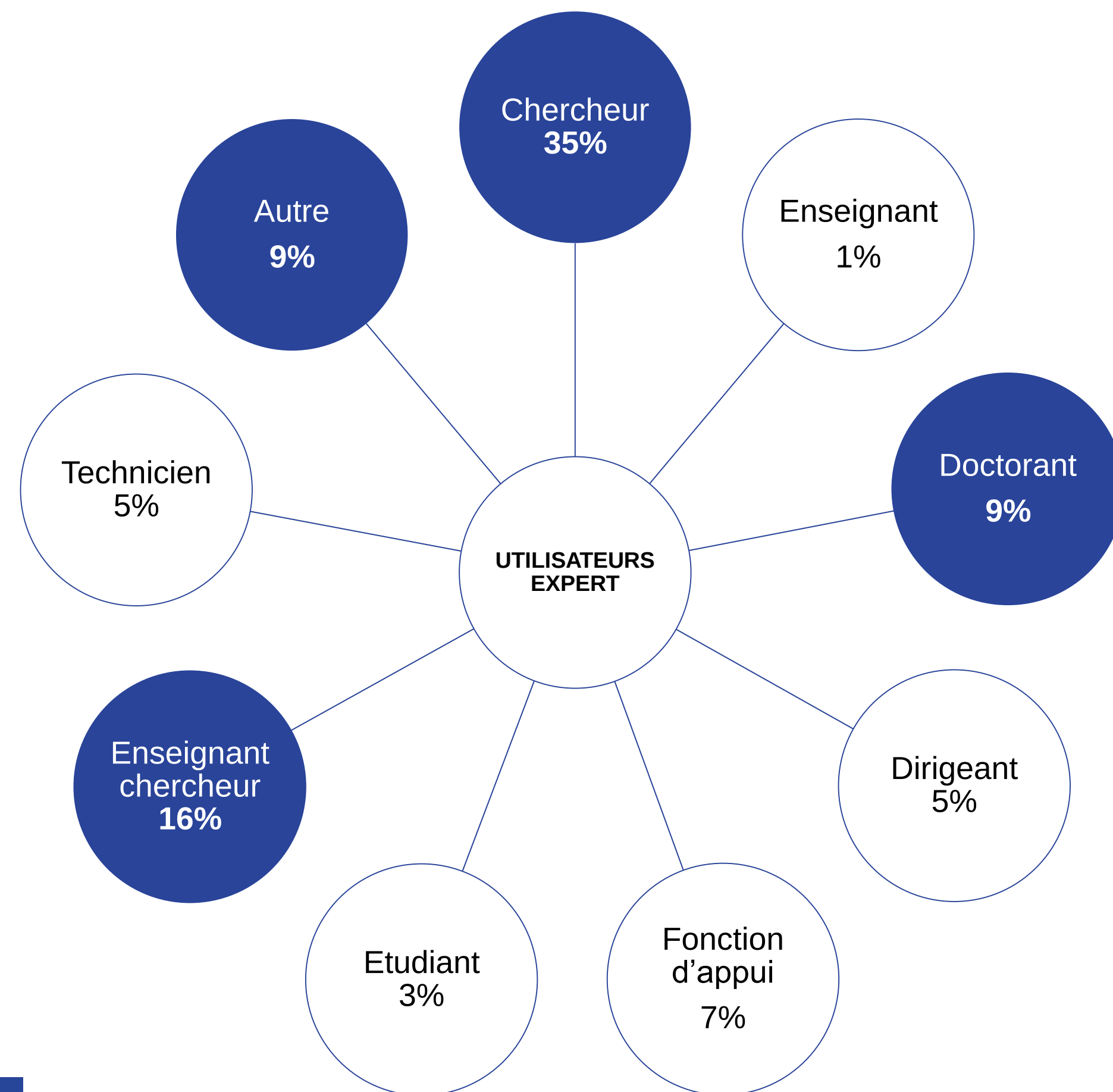
102 réponses

Oui, en tant qu'utilisateur ponctuel

102 réponses



Pratique de la fouille de texte / utilisateur expert

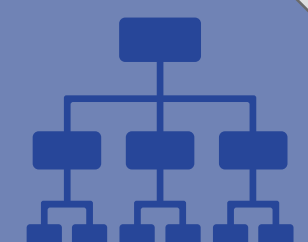


75 réponses

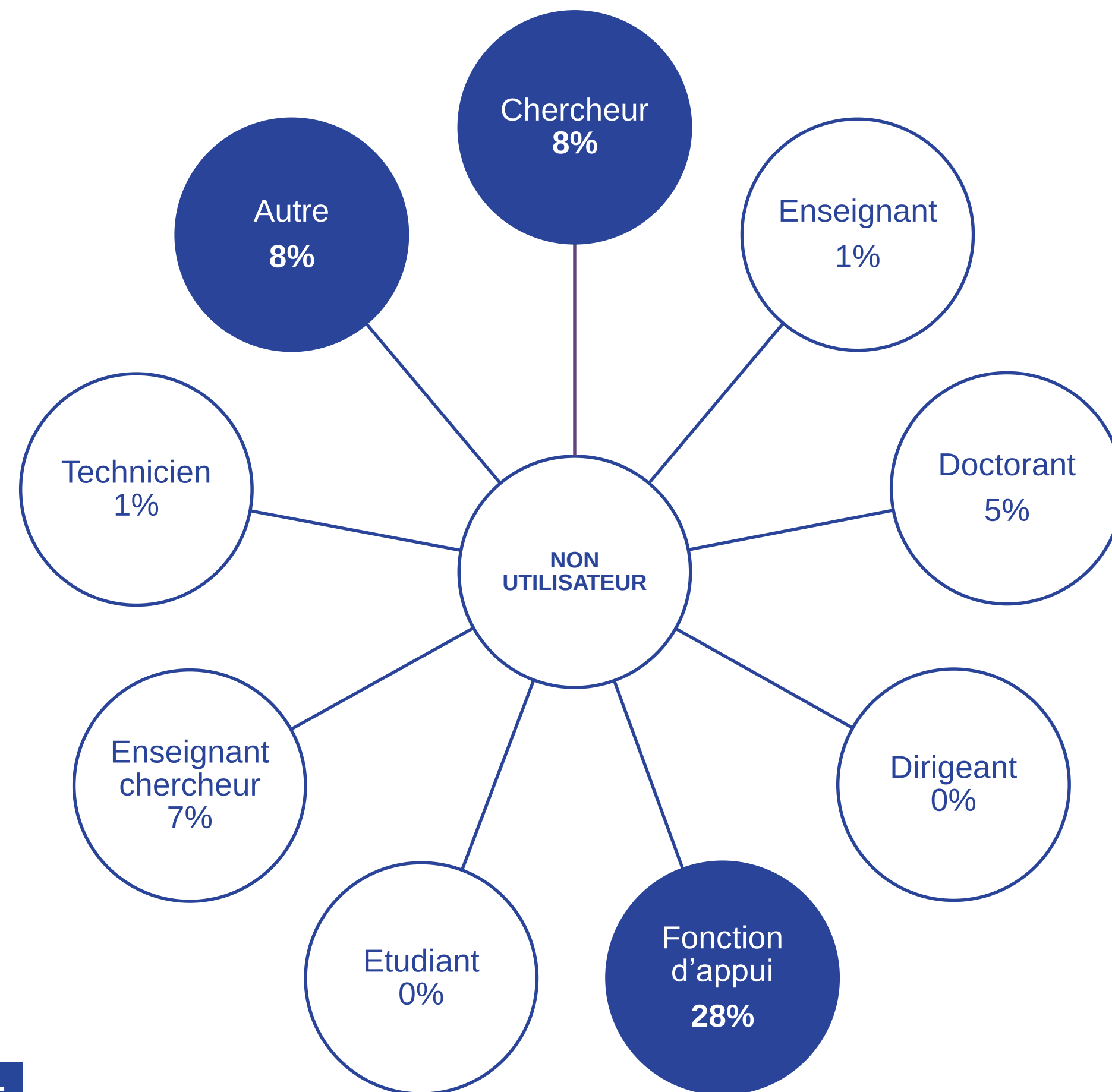
Oui, en tant qu'expert

AUTRE

Chef de projet, ingénieur d'étude, ingénieur de recherche, ingénieur data, documentaliste, consultant département R&D, chargé de projet dans les outils numériques, architecte, développeur



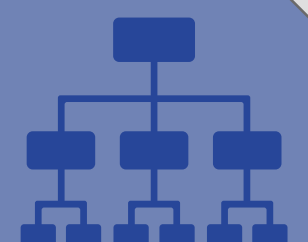
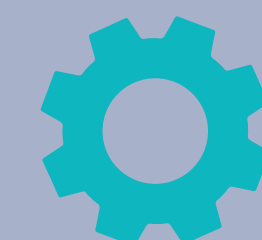
Pratique de la fouille de texte / non utilisateur



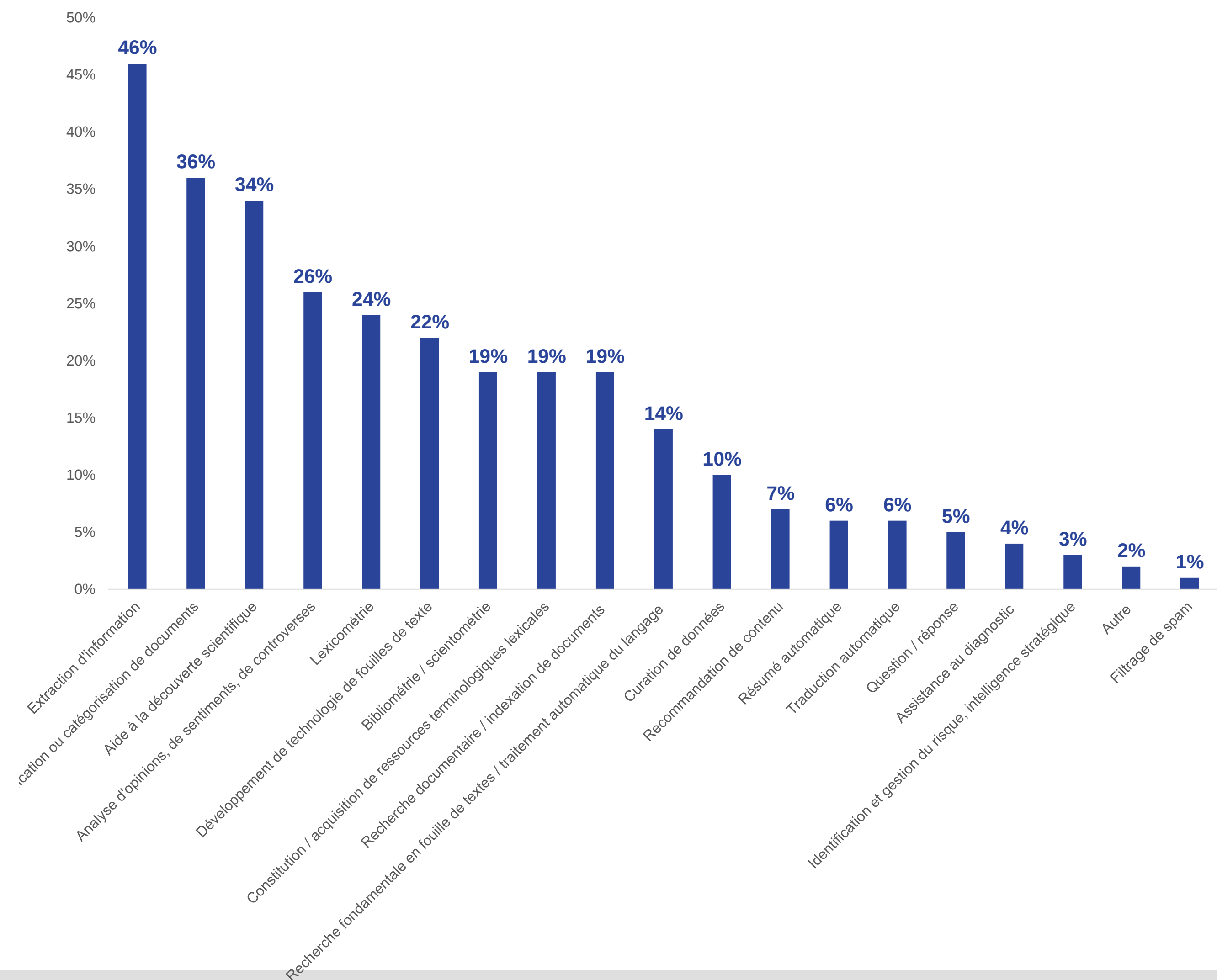
36 réponses
Non utilisateur

AUTRE

Chef de projet, ingénieur d'étude, ingénieur de recherche, ingénieur data, documentaliste, consultant département R&D, chargé de projet dans les outils numériques, architecte, développeur



Précisez quelle utilisation vous faites de la fouilles de textes ?



46%

Extraction d'information

36%

Classification ou catégorisation de documents

34%

Aide à la découverte scientifique

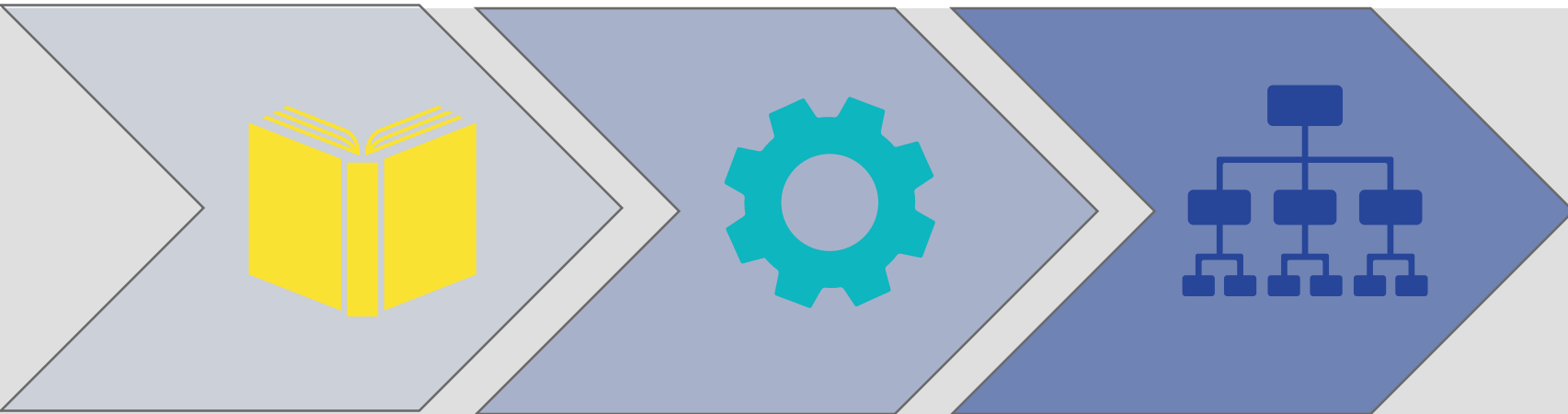
26%

Analyse d'opinions, de sentiments, de controverses

24%

Lexicométrie

178 réponses

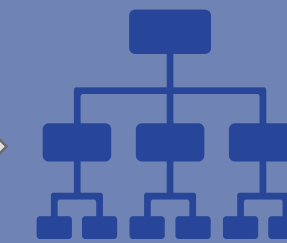
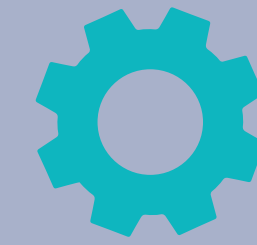


Précisez quelle utilisation vous faites de la fouilles de textes ? / profil des utilisateurs

Profil	Extraction d'information	Classification ou catégorisation de documents	Aide à la découverte scientifique	Analyse d'opinions, de sentiments, de controverses	Lexicométrie	Développement de technologie de fouille de texte	Bibliométrie / scientométrie	Constitution / acquisition de ressources terminologiques lexicales	Recherche documentaire / indexation de documents	Recherche fondamentale en fouille de textes
Chercheur	28%	23%	30%	34%	22%	32%	36%	28%	33%	42%
Dirigeant	3%	2%	4%	3%	2%	8%	2%	2%	2%	6%
Doctorant	7,5%	8%	12,5%	5%	10%	10%	7%	14%	7%	3%
Enseignant	19%	2%	4%	3%	4%	0%	2%	2%	2%	0%
Enseignant -chercheur	19%	19%	20%	19%	26,5%	12%	11%	23%	9%	10%
Etudiant	2%	4%	2,5%	3%	2%	0%	0%	0%	2%	0%
Fonction d'Appui	13%	16%	16%	10%	20%	6%	23%	16%	13%	13%
Technicien	7%	2%	4%	3%	10%	4%	2%	7%	4%	6%
Autre	7%	10%	2,5%	10%	10%	14%	4,5%	2%	11%	6%

AUTRE

Chef de projet, ingénieur d'étude, ingénieur de recherche, ingénieur data, documentaliste, consultant département R&D, chargé de projet dans les outils numériques, architecte, développeur



Si vous le souhaitez, donnez quelques précisions

« Donner des exemples »

« Outil utilisé pour un doctorat »

« Thèse sur l'extraction automatique des données expérimentales »

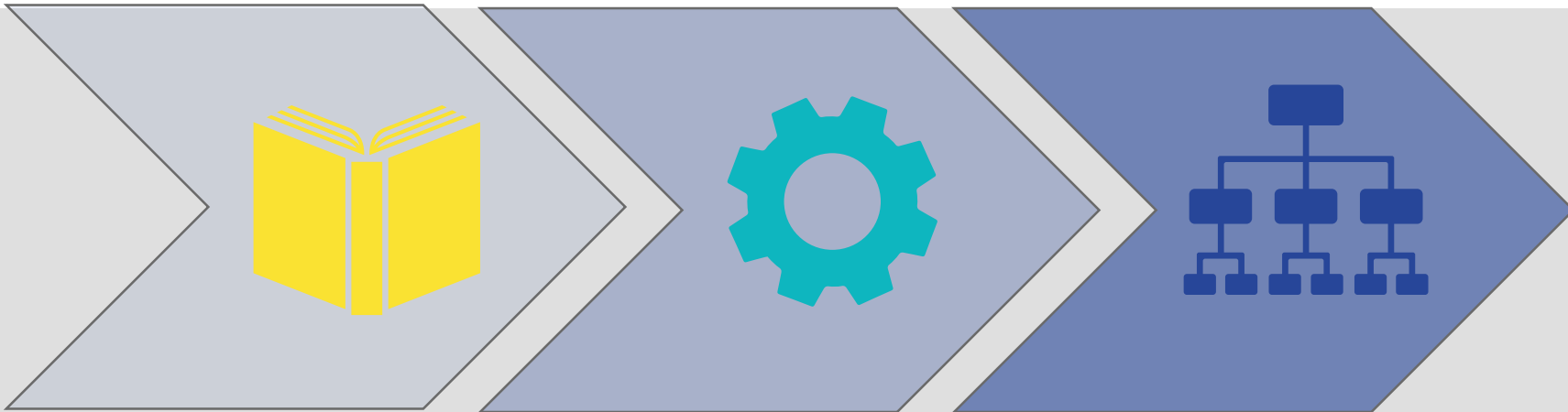
« Pour répondre à des questions de recherche »

« Projet de recherche fondamentale en fouille de textes » « Indexation sémantique »

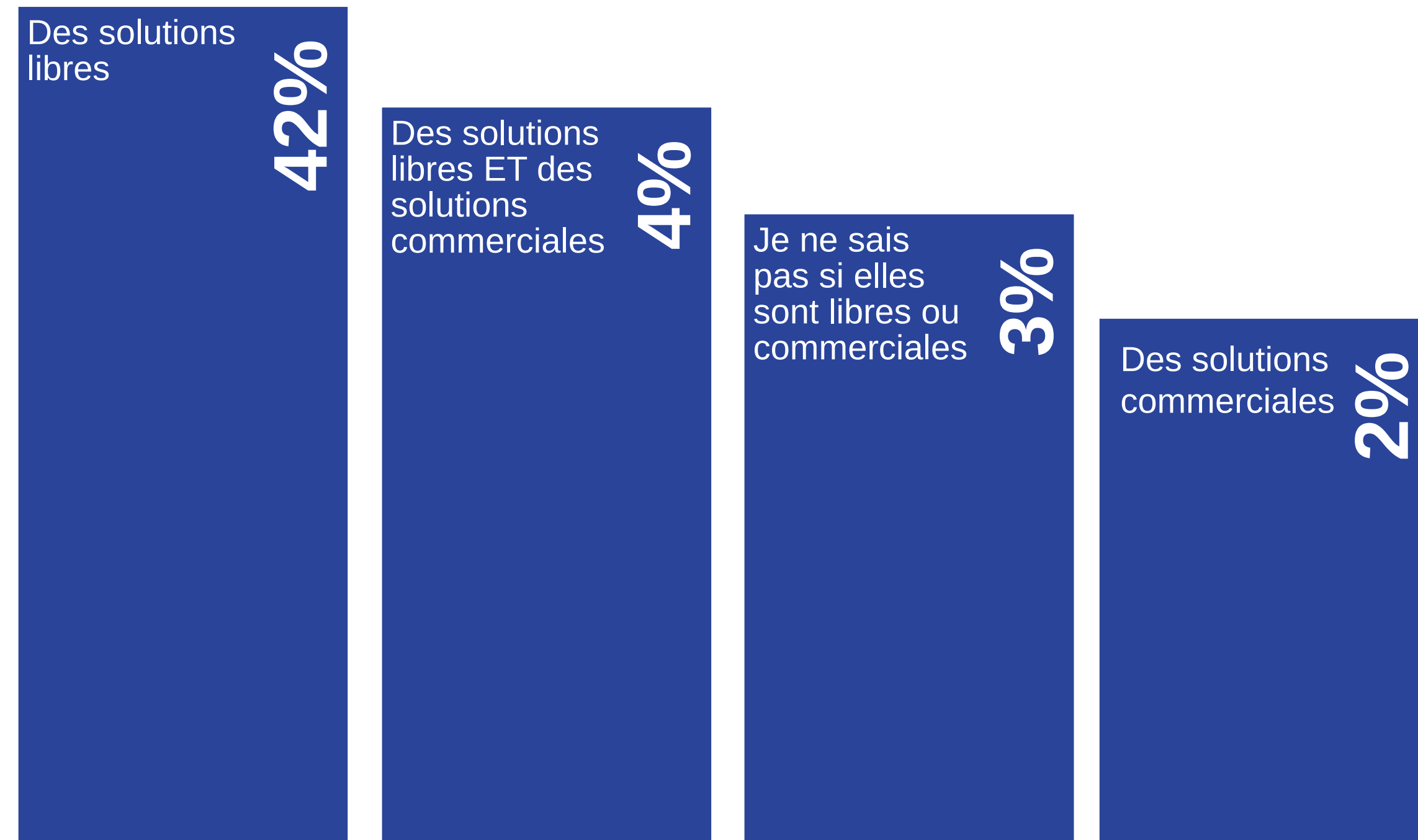
« Extraction de triggers »

« Analyse de publications scientifiques »

« Utilisation pour annotations sémantiques de données »



Quel type de technologies / suites logicielles dédiées à la fouille de texte utilisez-vous ?



42%

Des solutions libres

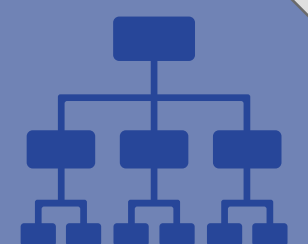
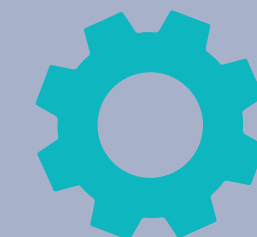
4%

Des solutions libres ET des solutions commerciales

3%

Je ne sais pas si elles sont libres ou commerciales

5 sans réponses



Pouvez-vous les citer ?

De A
à Z

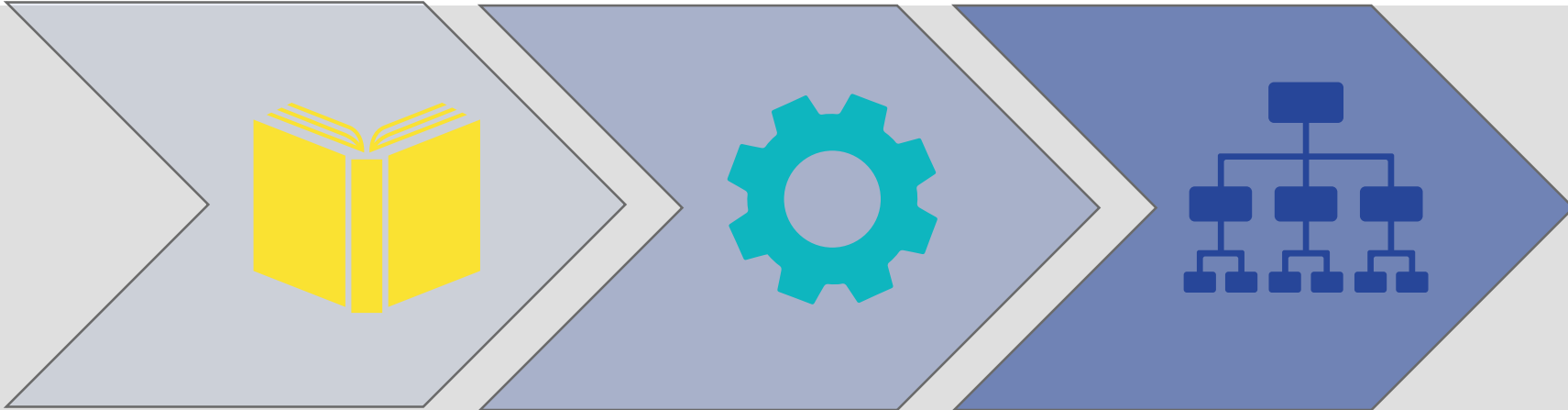
- Alceste
- Calliope
- Cogito
- CorText Manager
- Fasttext
- Gensim
- Gephy

- Google Scholar
- GroBid
- Iramuteq
- InVivo
- Knime
- Mathscinet

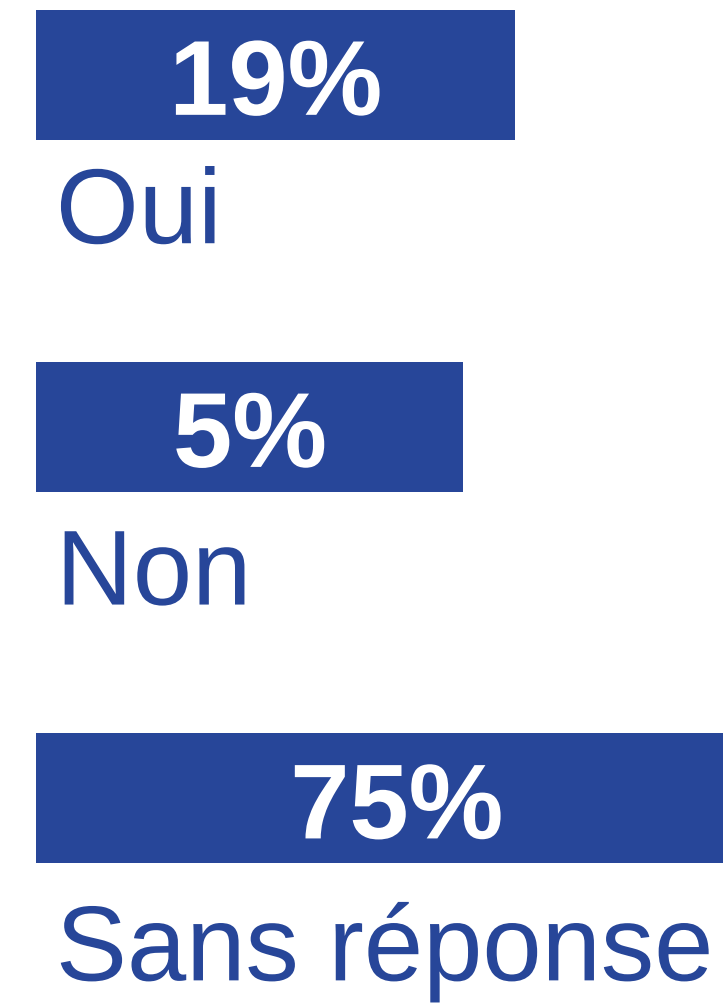
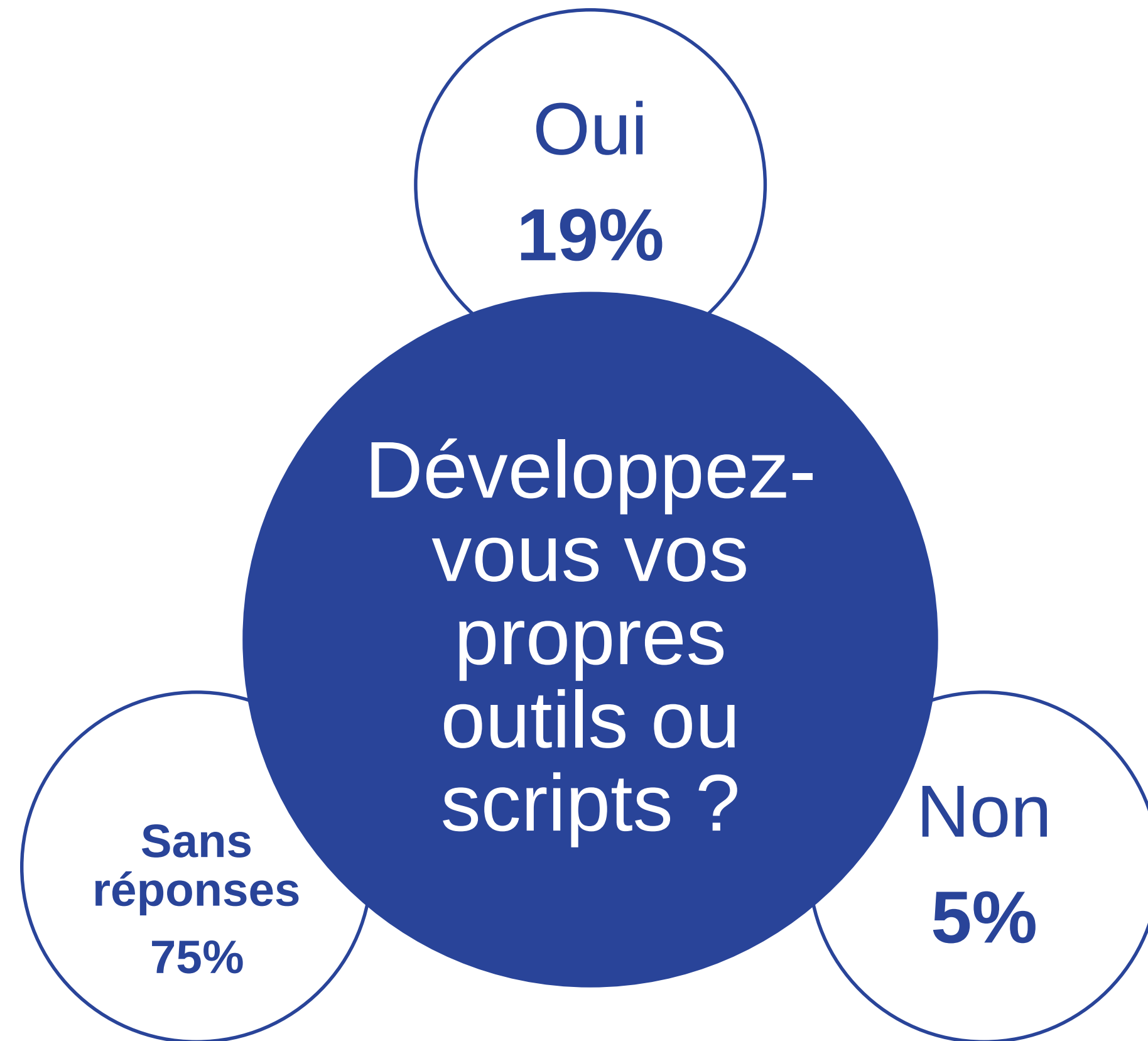
- NERD
- Numpy
- NKRL
- OpenRefine
- Python
- Qwant
- R

- Rapidminer
- Rstudio
- StanFord
- TensorFlow
- Term suite
- TXM
- UNITEX

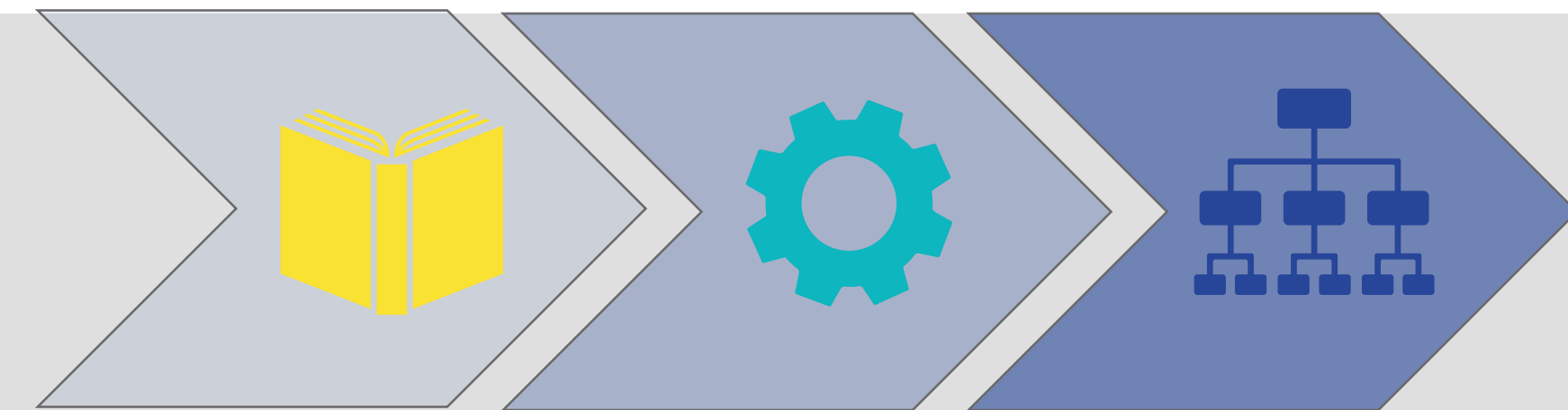
80 réponses



Développez-vous vos propres outils ou scripts pour la fouille de textes ?



73 réponses



Pouvez-vous nous indiquer lesquels en précisant leur type, une description, un lien vers une documentation ?

20
réponses

Alvis

Cartolab

Github

Google Scholar

GroBid

Pipeline Deep learning

Python

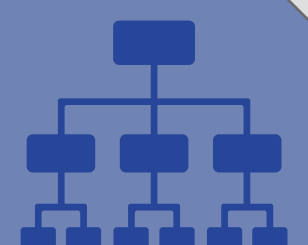
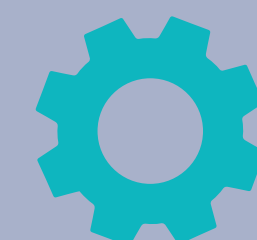
Semdee

TXM

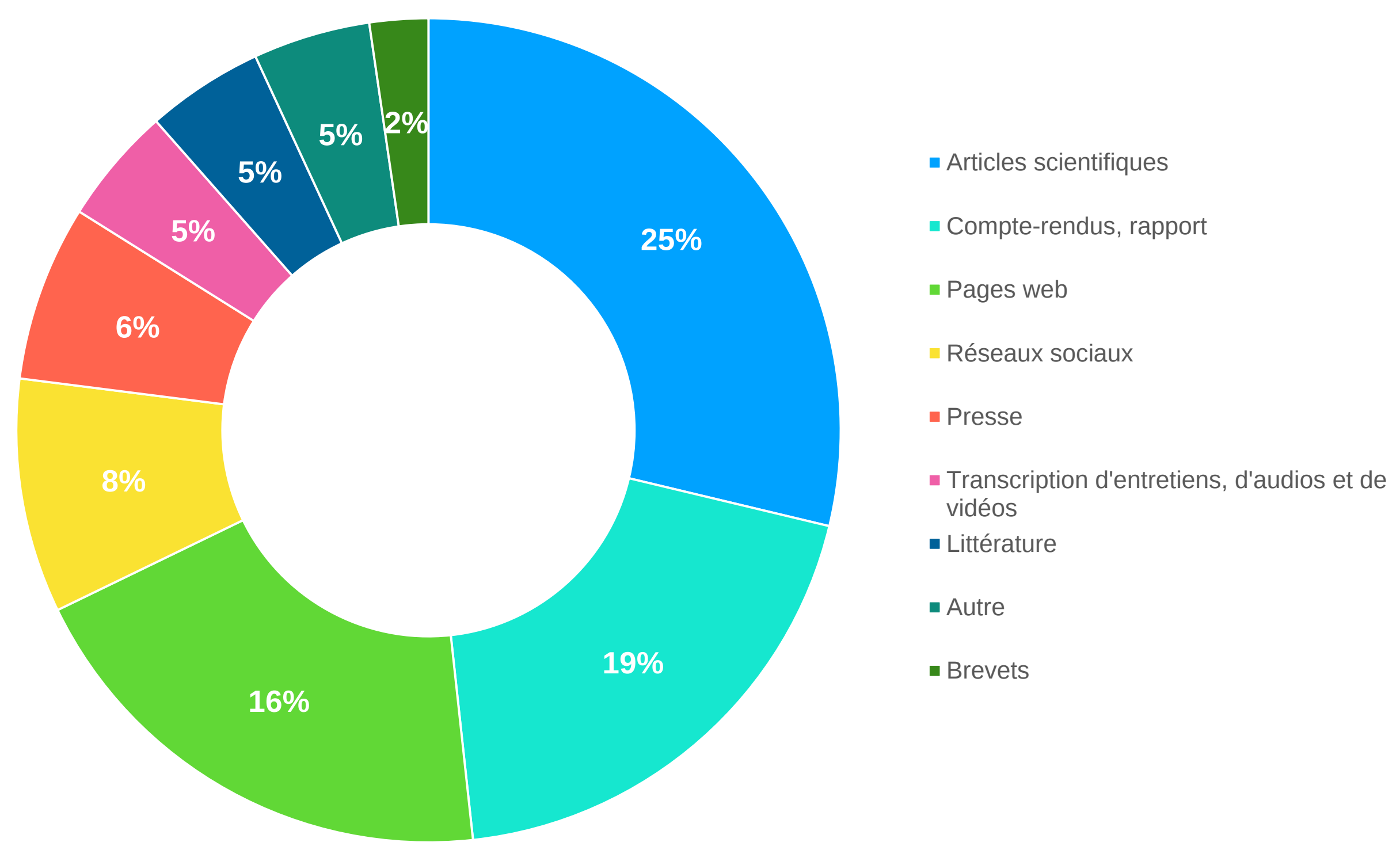
Unitex

LIEN

BioNLP-ST, Ontobiotope, ontologie WTO, BioPortal, PGxCorpus, Phare, PubAnnotation, Github, archives ouvertes, Renom, 80 jours, Melodi



Quels types de documents constituent le support de votre activité ?



25%
Articles scientifiques

19%
Pages web

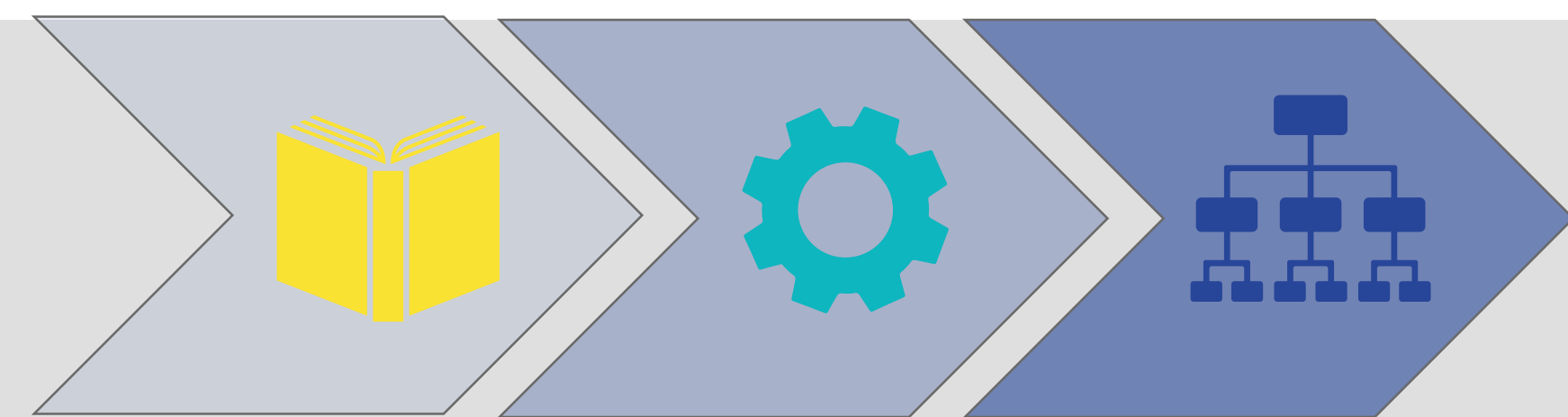
16%
Compte-rendu, rapport

8%
Réseaux sociaux

6%
Presse

AUTRE

archives, supports de formation, documents graphiques, revue des sociétés, métadonnées, jurisprudence

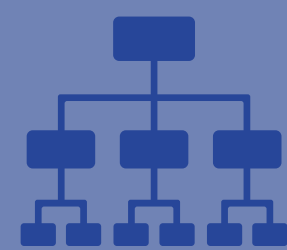
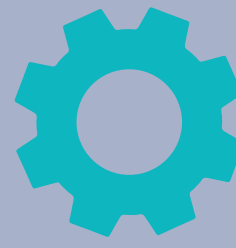


Quels types de documents constituent le support de votre activité ? / profil des utilisateurs

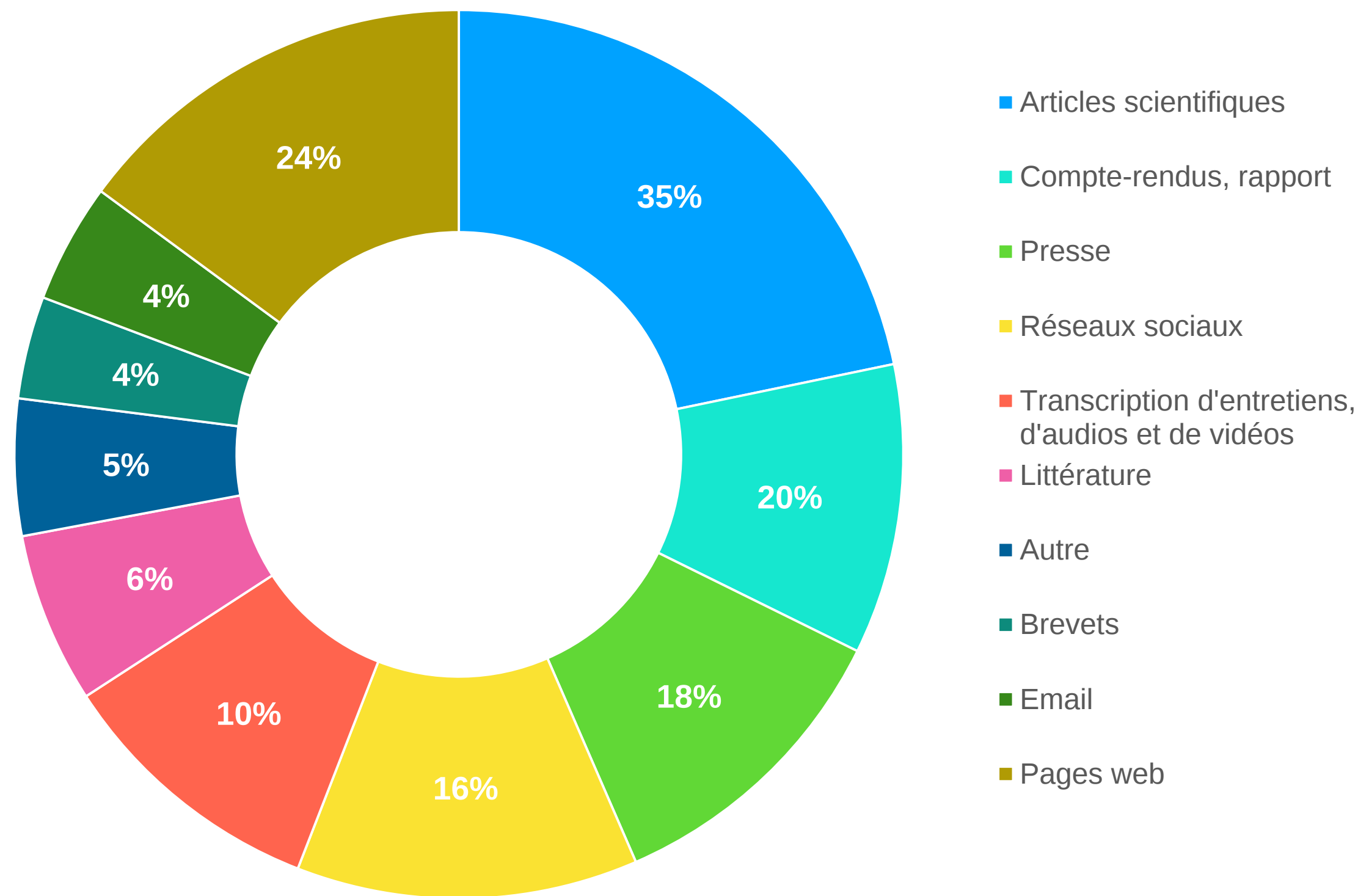
Profil	Articles scientifiques	Pages Web	Compte-rendu, rapport	Réseaux sociaux	Presse
Chercheur	10,5%	10%	12%	8%	10%
Dirigeant	0%	0%	0%	0%	0%
Doctorant	6,5%	6%	8%	4%	5%
Enseignant	1%	2%	2%	0%	0%
Enseignant-chercheur	9%	9%	10%	8%	15%
Etudiant	0%	0%	0%	0%	0%
Fonction d'Appui	28%	33%	28%	21%	35%
Technicien	1%	2%	2%	4%	0%
Autre	8%	12%	12%	21%	15%

AUTRE

archives, supports de formations, documents graphiques, revues des sociétés, métadonnées, jurisprudence



Parmi ces types de documents, sur lesquels appliquez-vous la fouille de textes ?



35%

Articles scientifiques

24%

Pages Web

20%

Réseaux sociaux

18%

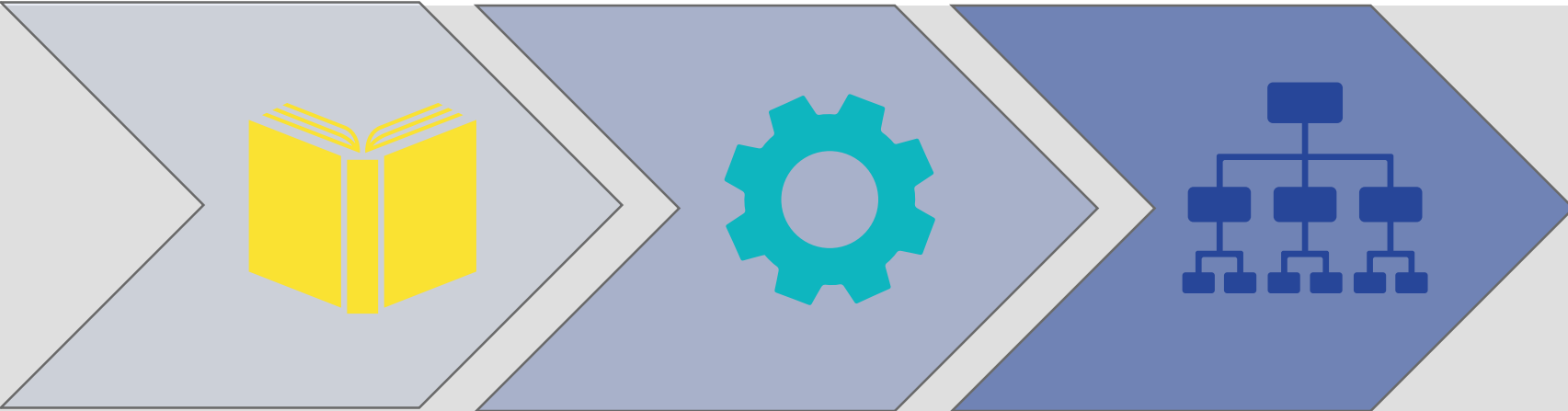
Presse

16%

Transcription d'entretiens, d'audios, vidéos

AUTRE

archives, corpus, notices bibliographiques, jurisprudence, discours, BDD, avis clients

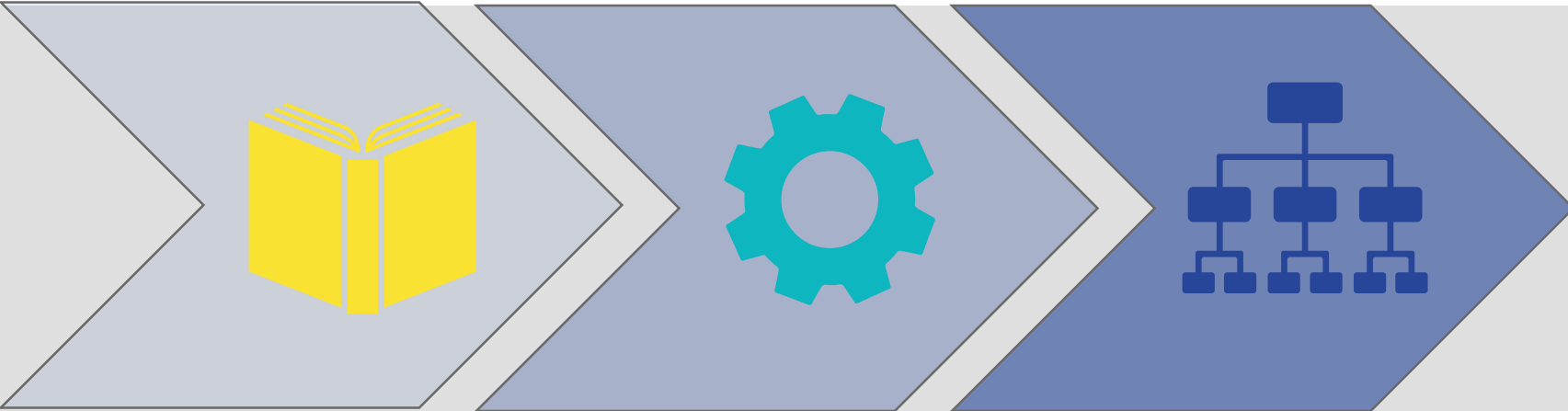


Parmi ces types de documents, sur lesquels appliquez-vous la fouille de textes ? Profil des utilisateurs

Profil	Articles scientifiques	Pages Web	Réseaux sociaux	Presse	Transcription d'entretiens, d'audios, de vidéos
Chercheur	29%	30,5%	32%	22%	21%
Dirigeant	5%	7%	7%	5,5%	6%
Doctorant	9%	8%	5%	9%	10%
Enseignant	4%	4%	3%	4%	4%
Enseignant-chercheur	14%	19%	27%	28%	23%
Etudiant	1%	1%	2%	2%	0%
Fonction d'Appui	17%	8%	8%	5,5%	12,5%
Technicien	5%	5,5%	5%	4%	12,5%
Autre	8%	12,5%	8%	9%	8%

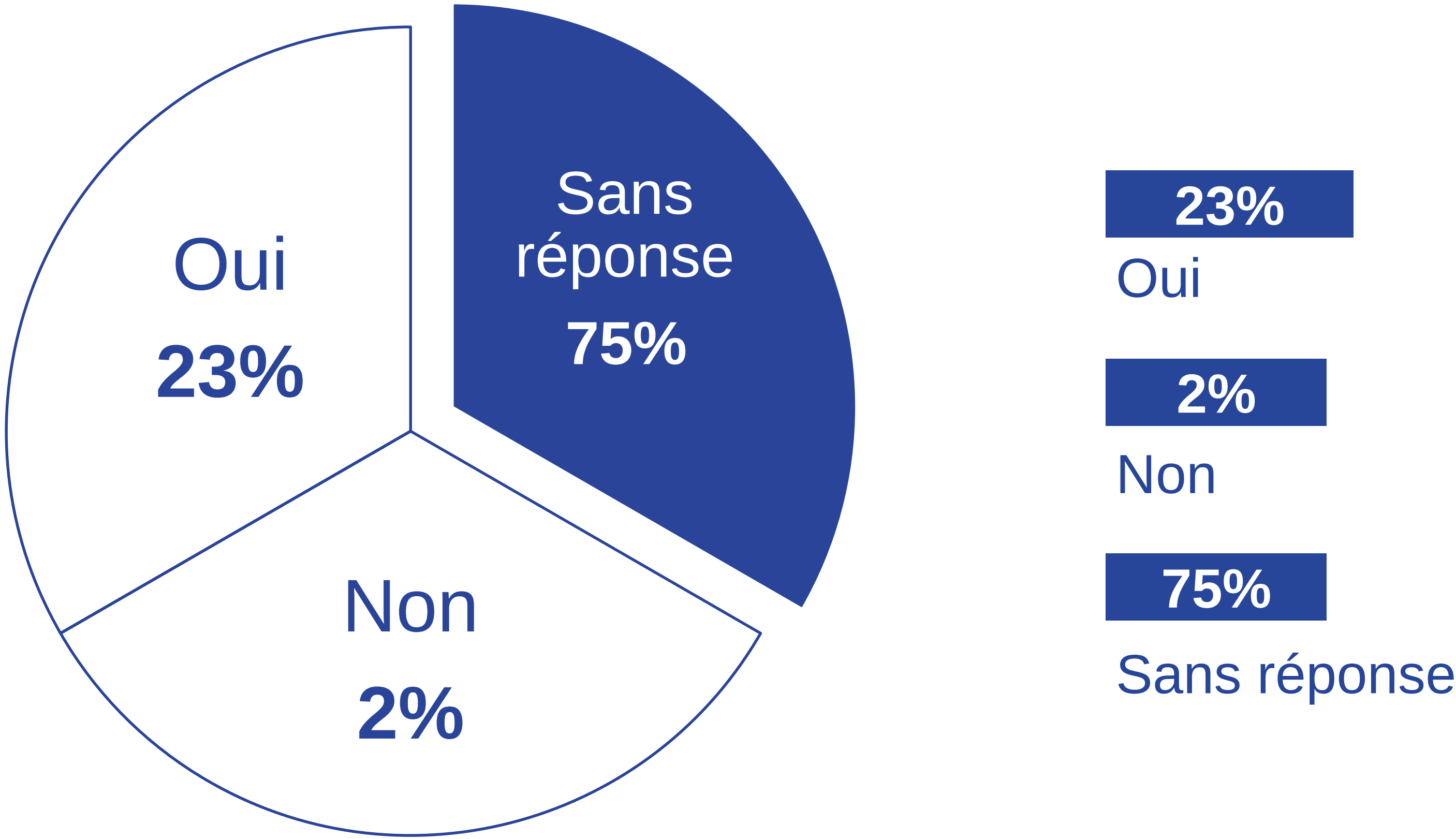
AUTRE

archives, corpus, notices bibliographiques, jurisprudence, discours, BDD, avis clients

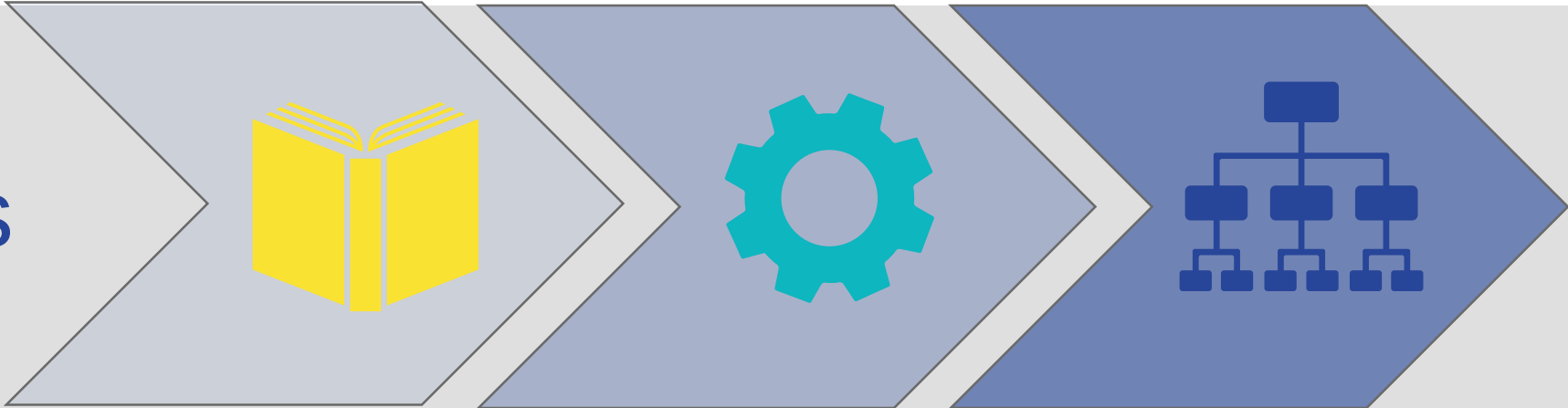


Vous arrive-t-il d'appliquer des pré-traitements à vos corpus préalablement à la fouille de textes ?

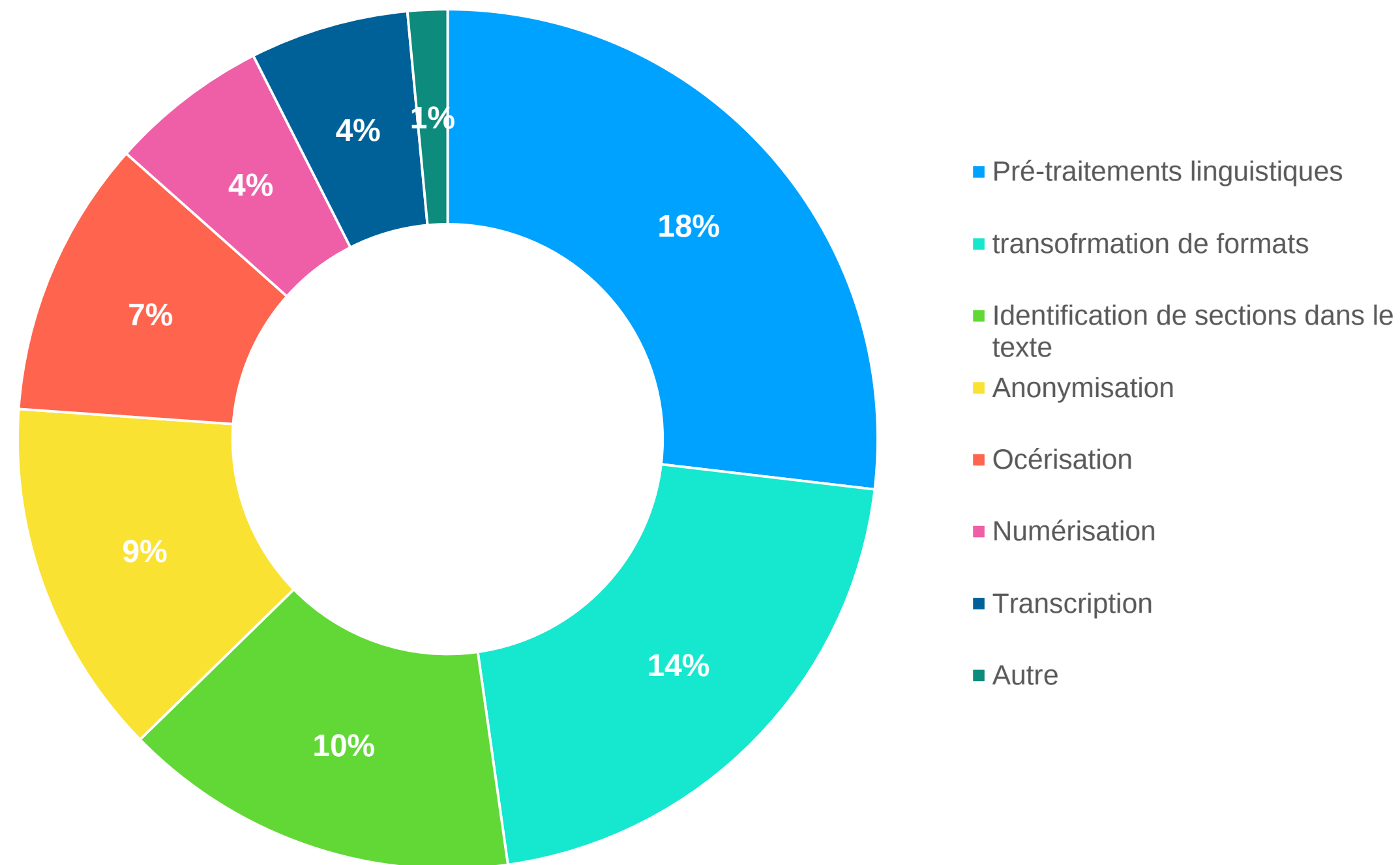
74
réponses



Les utilisateurs ayant répondu **OUI** sont 100% des utilisateurs experts



De quels types sont ces pré-traitements ?



18%

Pré-traitements linguistiques

14%

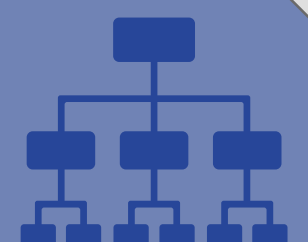
Transformation de formats

10%

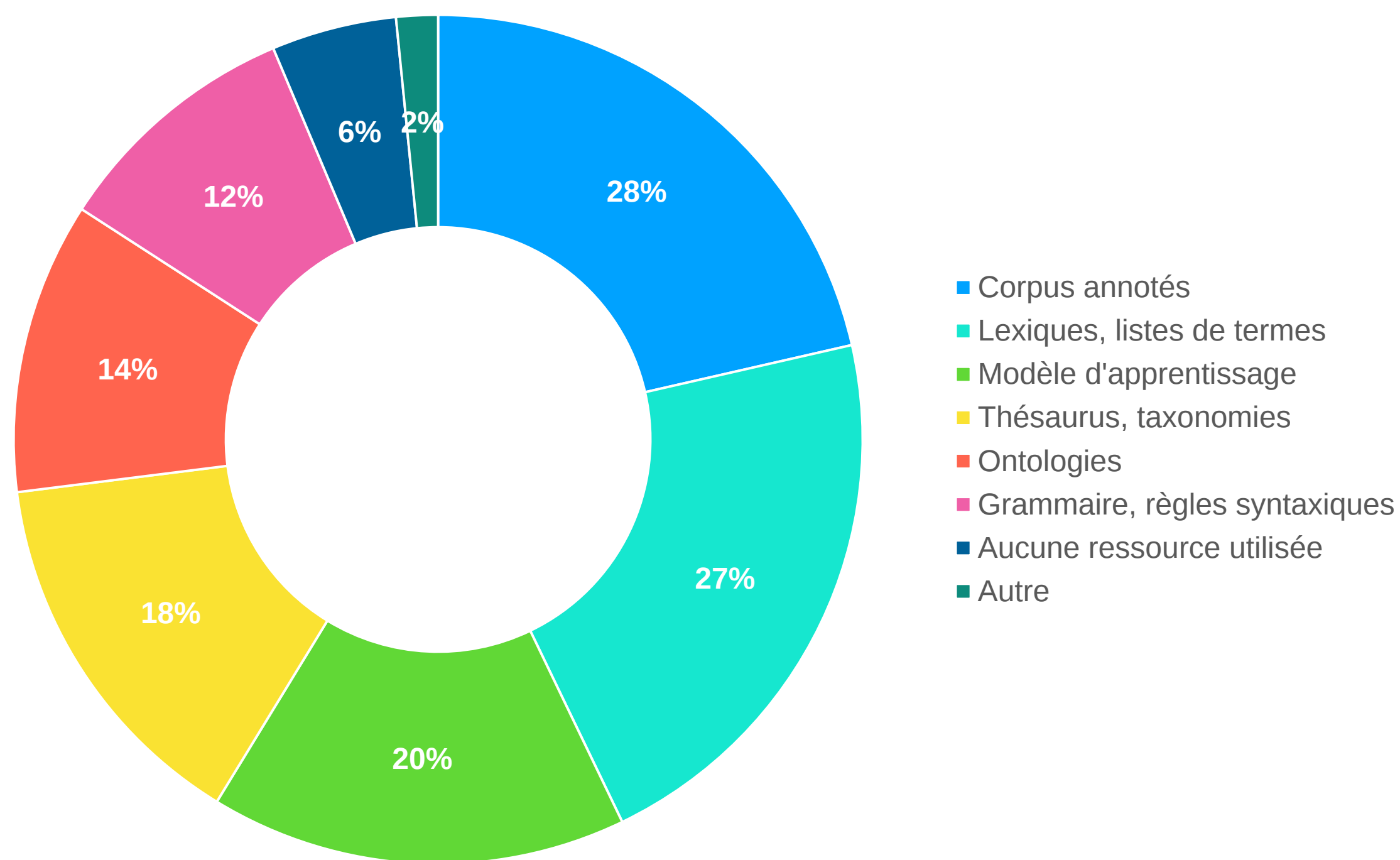
Identification de sections dans le texte

AUTRE

ré-encodage, statistiques, ajout de variables



Quels types de ressources utilisez-vous pour vos traitements de text-mining?



28%
Lexiques, listes de termes

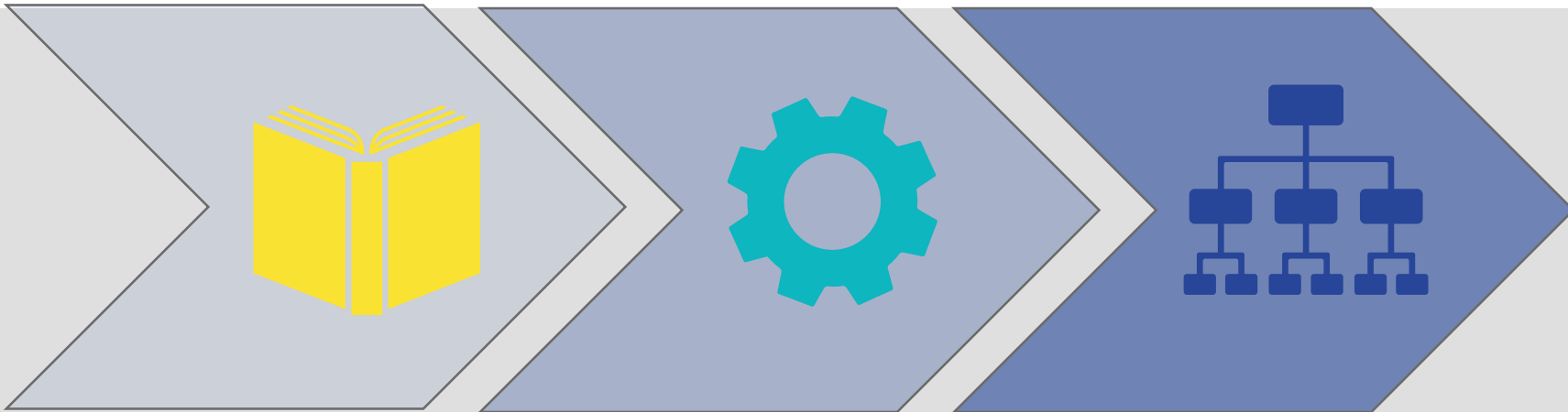
27%
Corpus annotés

20%
Modèles d'apprentissage

18%
Thésaurus, taxonomies

AUTRE

word embedding, corpus non annotés, analyse des graphes

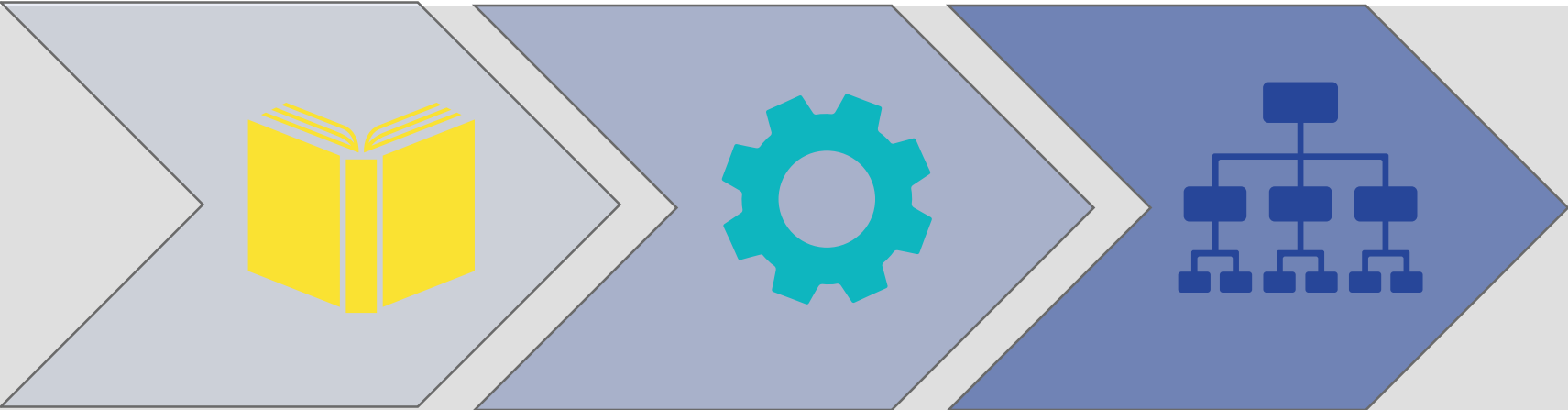


Quels types de ressources utilisez-vous pour vos traitements de text-mining? / profil des utilisateurs

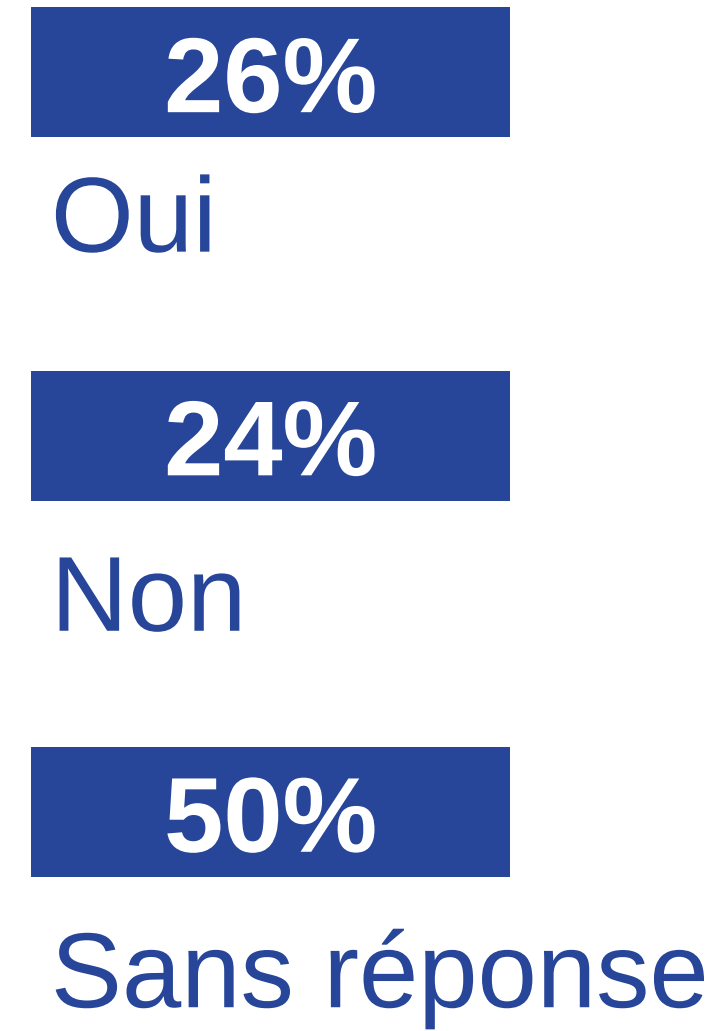
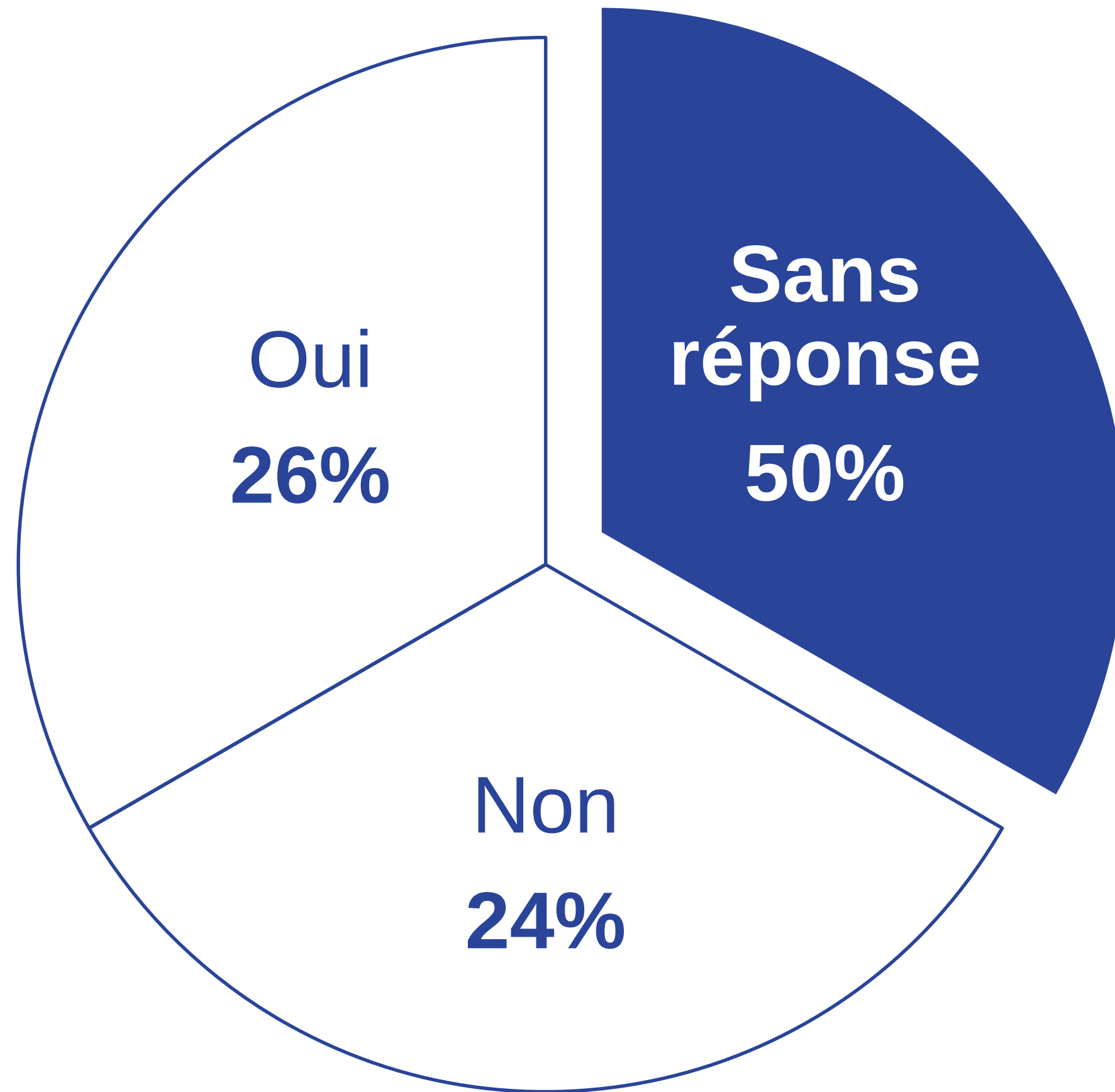
Profil	Lexiques, listes de termes	Corpus annotés	Modèles d'apprentissage	Thésaurus, taxonomies	Ontologies
Chercheur	30%	26%	33%	17%	23%
Dirigeant	2%	0%	3%	3%	4,5%
Doctorant	2%	6%	3%	3%	4,5%
Enseignant	0%	0%	0%	0%	0%
Enseignant-chercheur	13%	9%	7%	17%	18%
Etudiant	0%	0%	0%	0%	0%
Fonction d'Appui	19,5%	18%	27%	10%	18%
Technicien	6,5%	6%	3%	7%	4,5%
Autre	11%	6%	7%	7%	9%

AUTRE

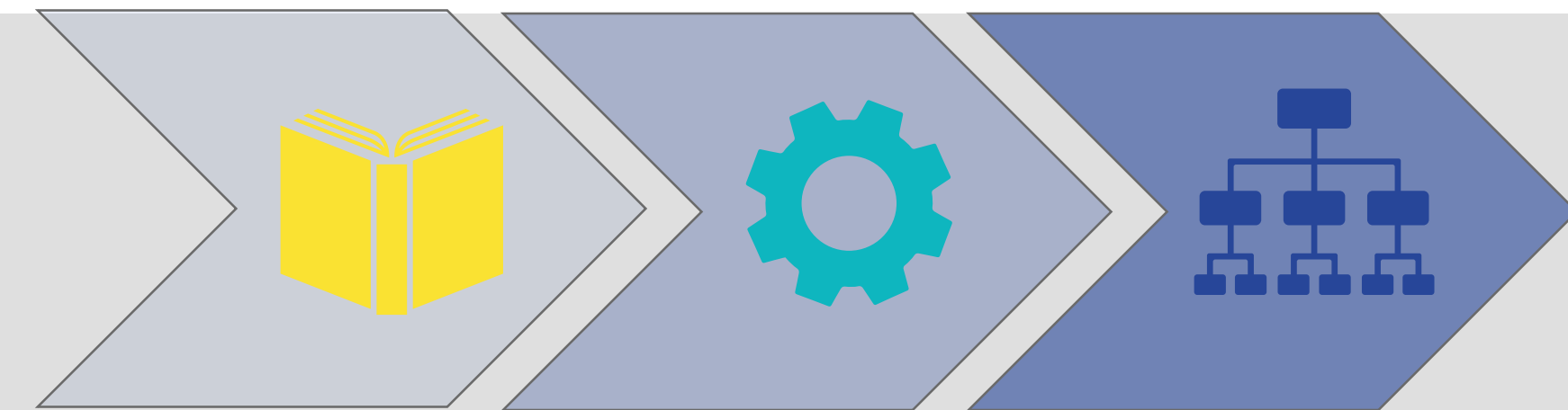
word embedding, corpus non annotés, analyse des graphes



Développez-vous vos propres ressources pour le text mining ?

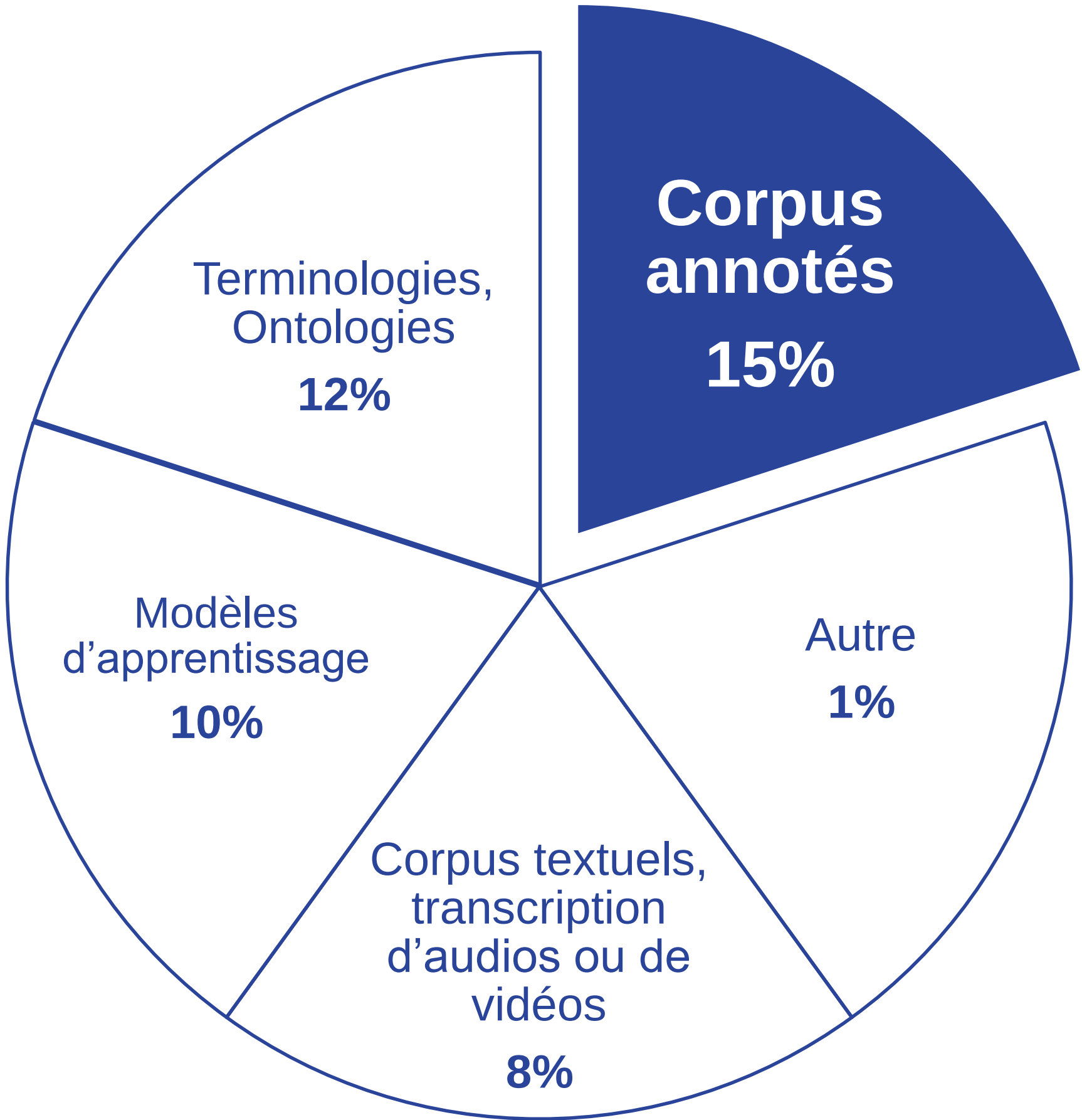


148 réponses



Que produisez-vous ?

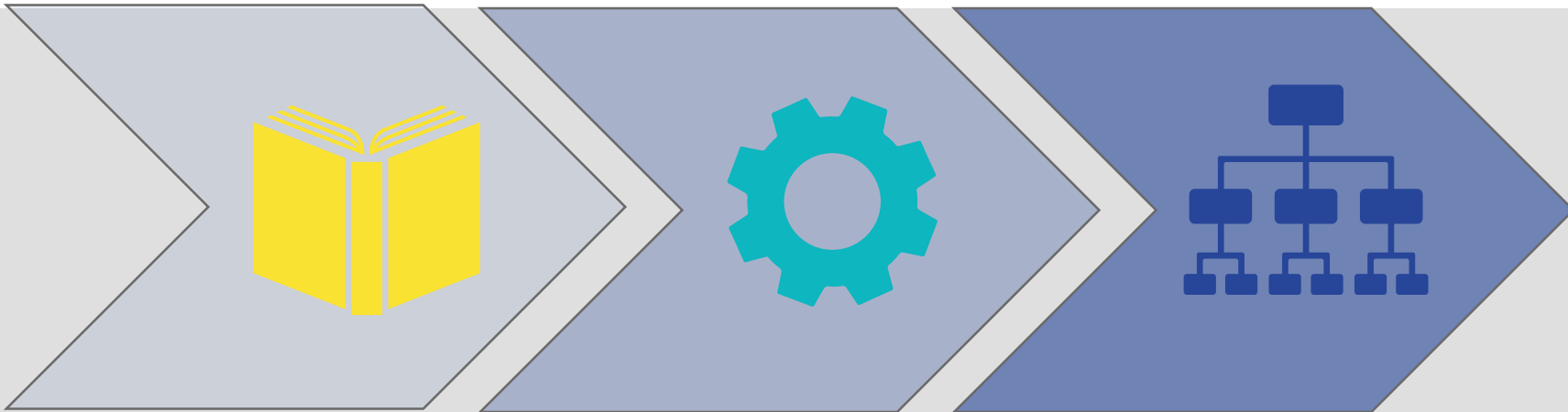
Les chercheurs et enseignants chercheurs produisent des modèles d'apprentissage



- 15%
Corpus annotés
- 12%
Terminologies, ontologies
- 10%
Modèles d'apprentissage

AUTRE

script, traitement, word embedding, modèles d'analyse, bases en langage formel



Pouvez-vous nous indiquer lesquels en précisant leur type, une description, un lien vers une documentation ?

De A
à Z

BFM

BioPortal

Corpus d'opinion

Cortypo

FrenchCropUsage

Github

Irit

ISTEX

My Web intelligence

Ontobiotope

Ontologie WTO

PGxCorpus

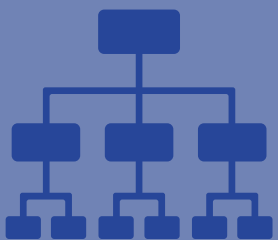
Prolexbase

Pub annotation

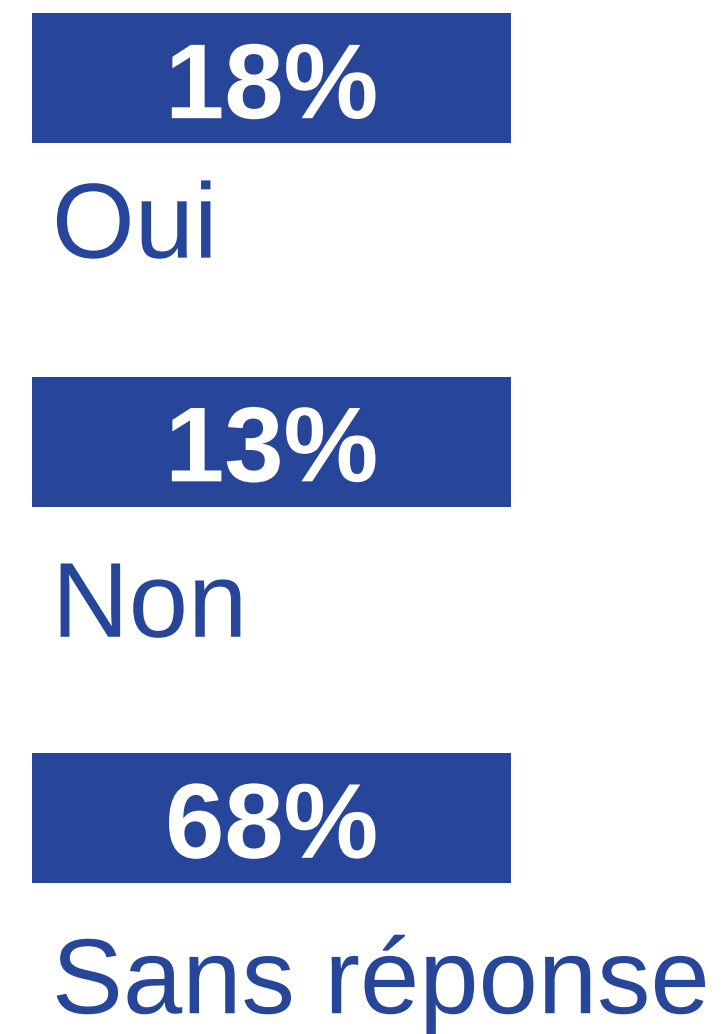
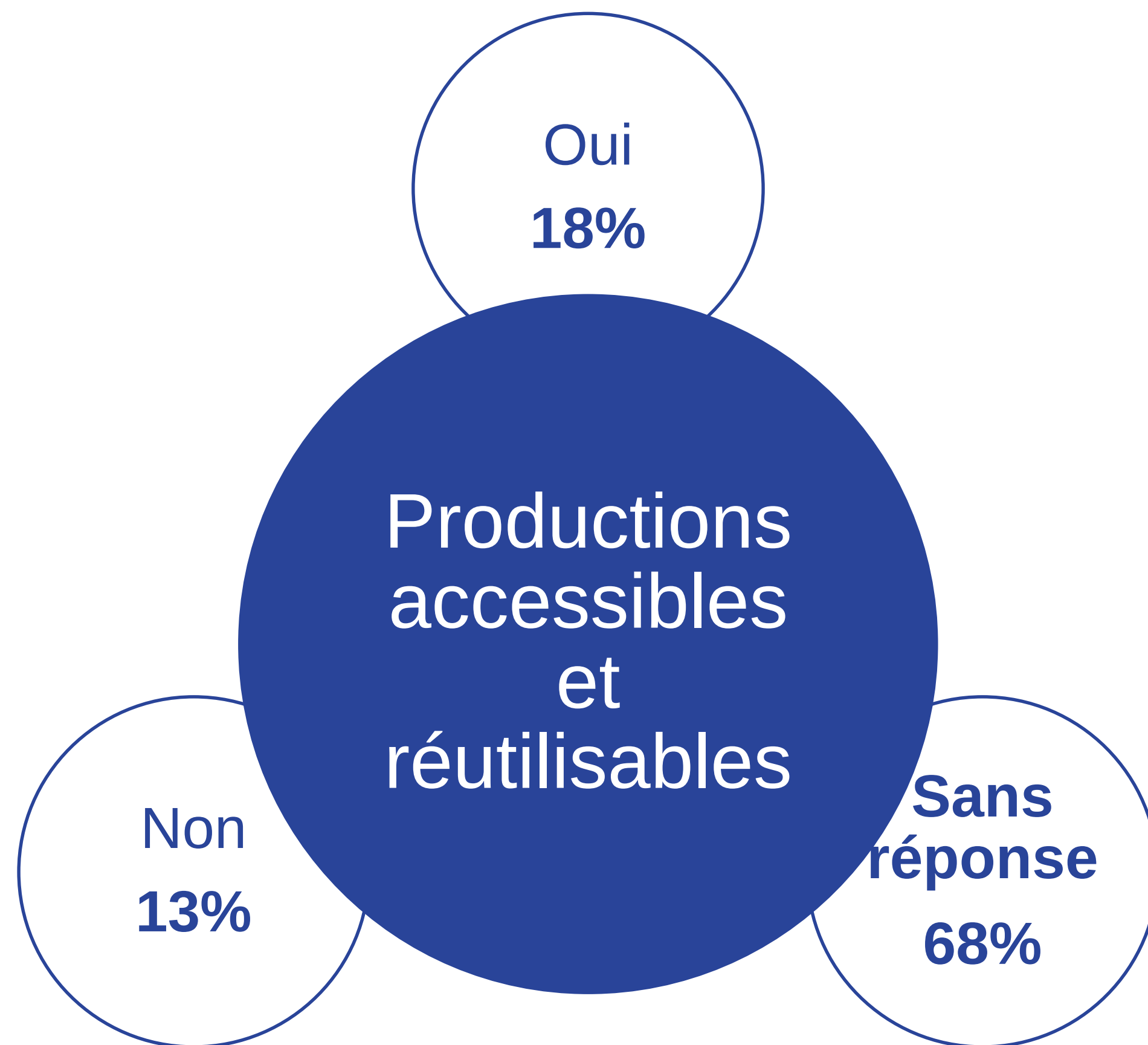
Publications libres
de droit

Lien

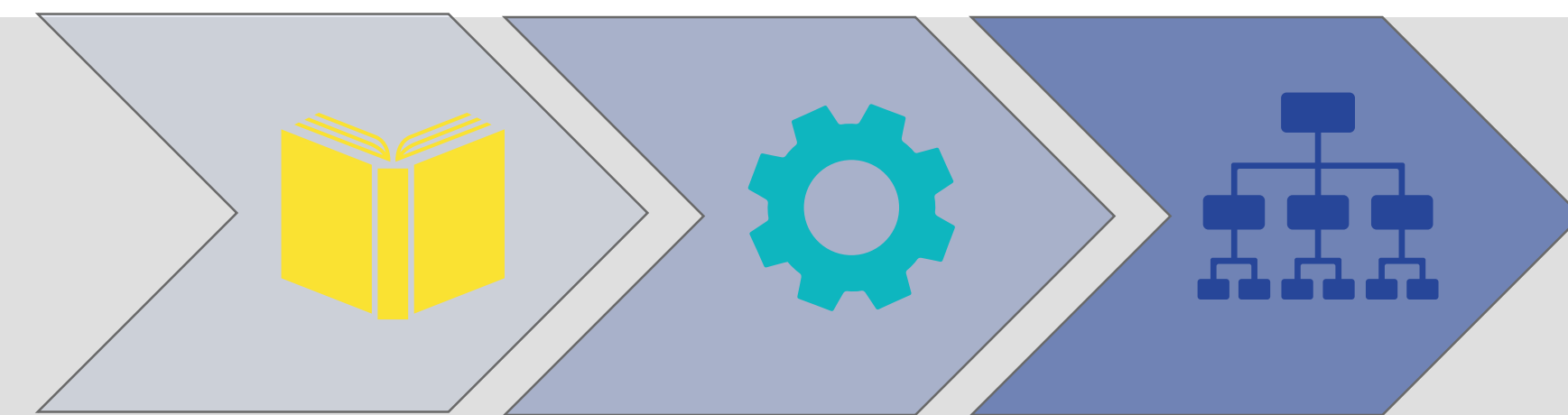
BFM, TreeTagger, BioNLP-ST, OntoBiotope, WTO, Bioportal, PGxCorpus, PHARE, Pubannotation, Melodi



Avez-vous l'habitude de rendre ces productions accessibles et réutilisables par d'autres ?

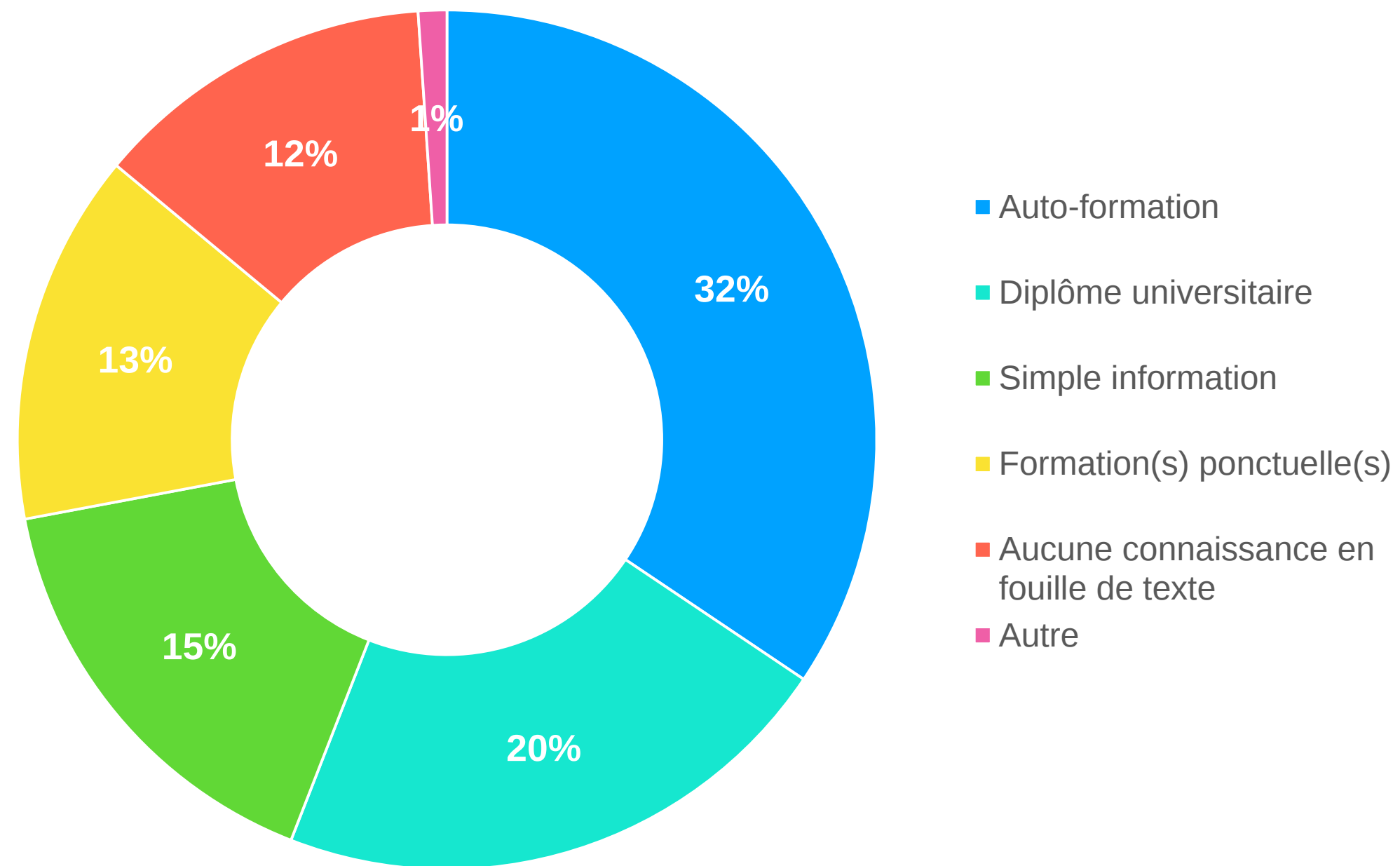


95 réponses



Compétences et formation

Quel est votre niveau de formation à la fouille de textes ?



32%

Auto-formation

20%

Diplôme universitaire

15%

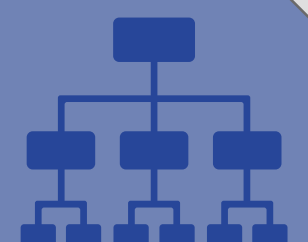
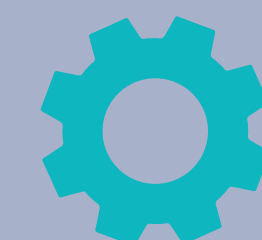
Simple information

13%

Formation(s) ponctuelle(s)

AUTRE

MOOC, formation complémentaire, recherche d'informations sur le web

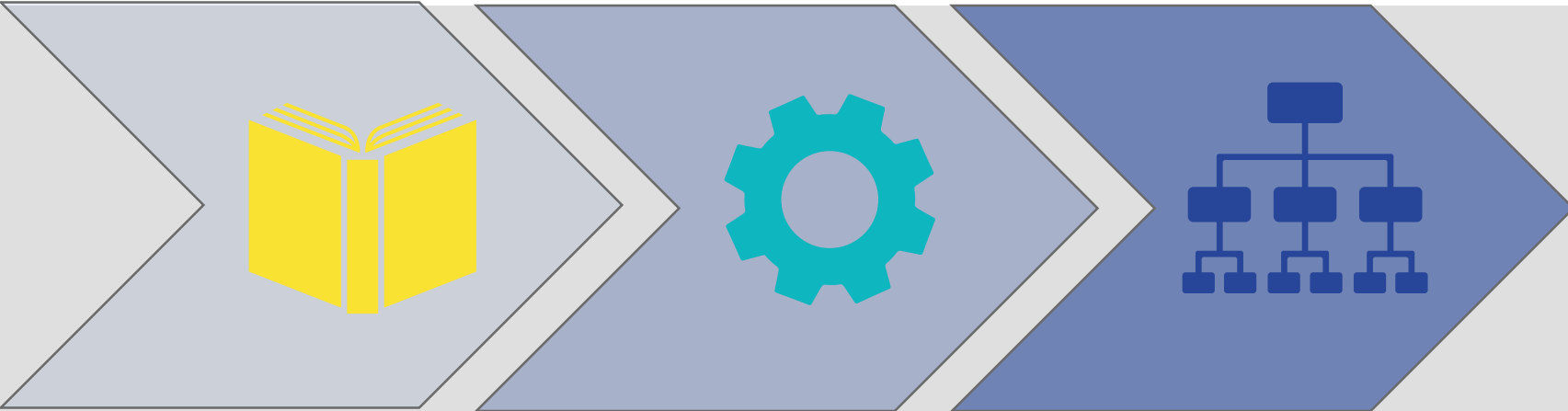


Quel est votre niveau de formation à la fouille de textes ? / profil des utilisateurs

Profil	Auto-Formation	Diplôme universitaire	Simple information	Formation(s) ponctuelle(s)
Chercheur	25%	33%	15,5%	31%
Dirigeant	3%	3%	0%	2,5%
Doctorant	10%	10%	4%	10%
Enseignant	4%	2%	2%	5%
Enseignant-chercheur	16%	16%	11%	5%
Etudiant	1%	5%	0%	5%
Fonction d'Appui	21%	3%	38%	31%
Technicien	6%	6,5%	4%	10%
Autre	7%	11%	4%	8%

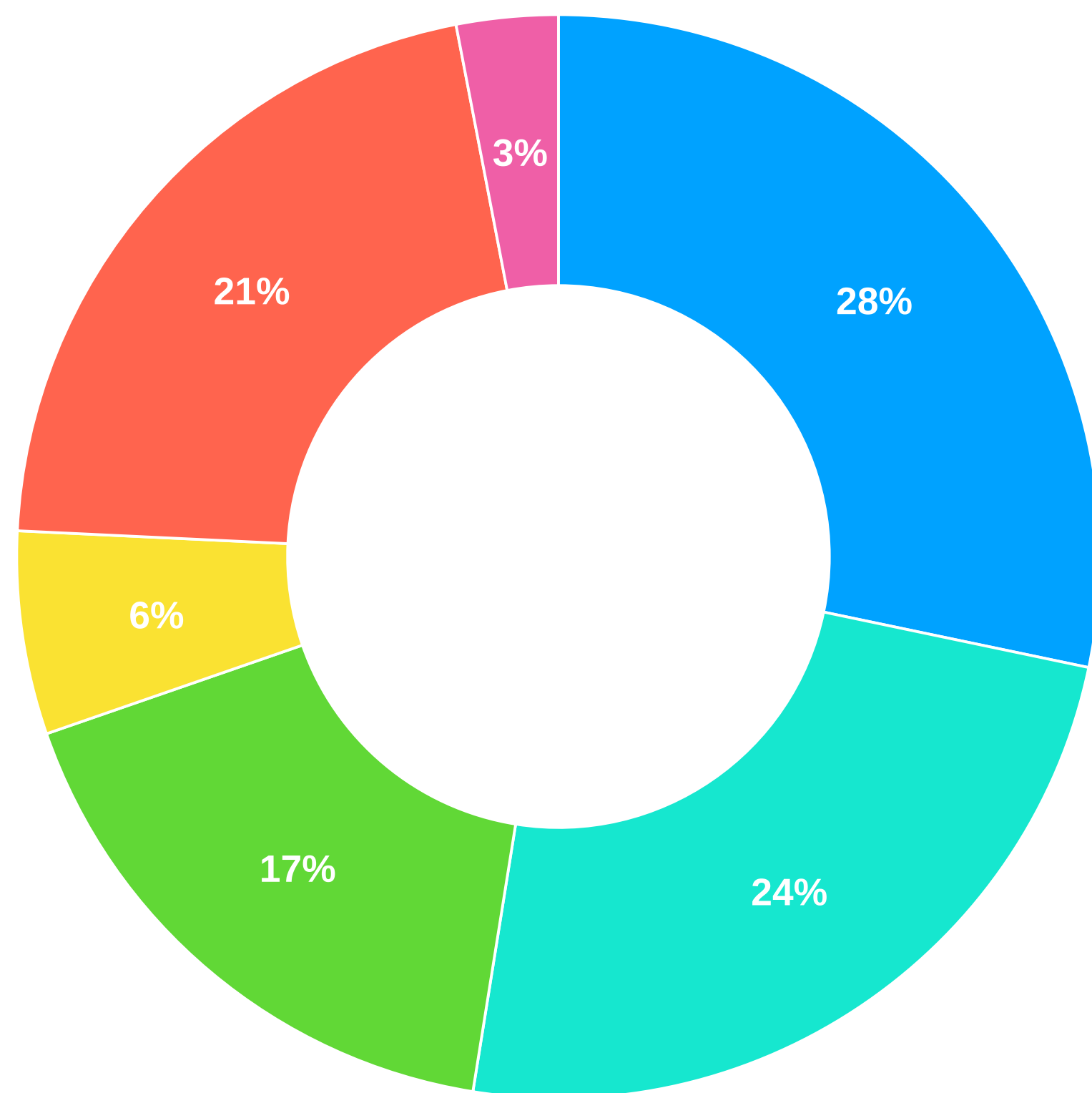
AUTRE

MOOC, formation complémentaire, recherche d'informations sur le web



Estimez-vous connaître vos droits en matière de fouille de texte ?

21%
Sans réponse



- Plutôt bien mais avec des hésitations
- J'en ai entendu parler
- Pas du tout
- Je ne m'en préoccupe pas
- Sans réponse
- Parfaitement bien

28%

Plutôt bien mais avec des hésitations

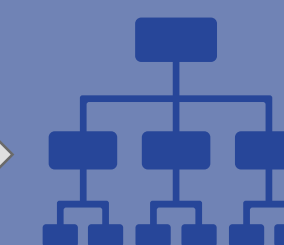
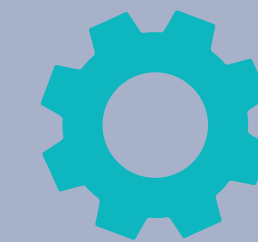
24%

J'en ai entendu parler

17%

Pas du tout

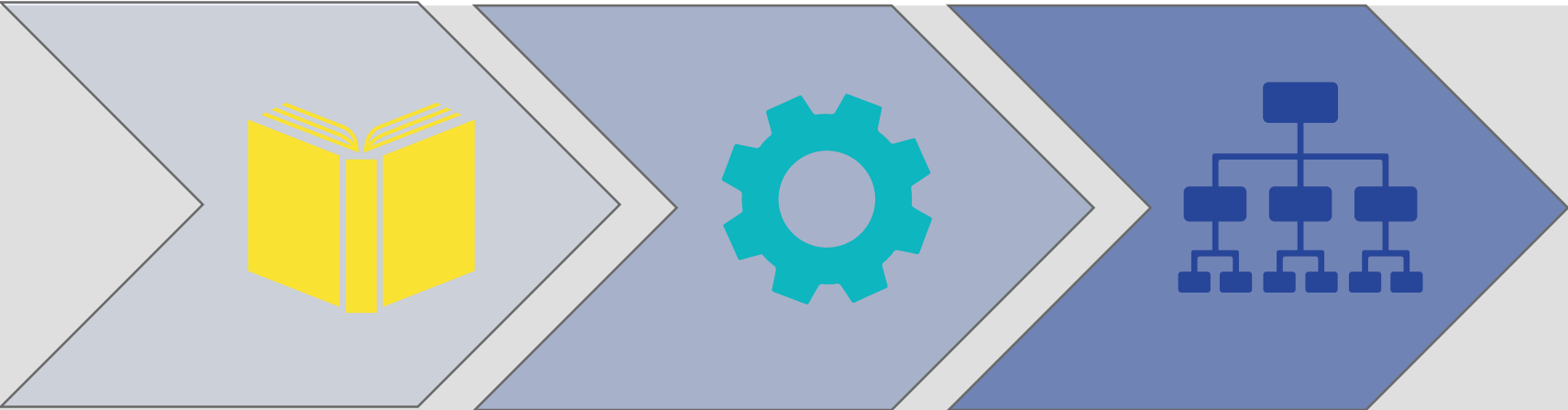
235 réponses



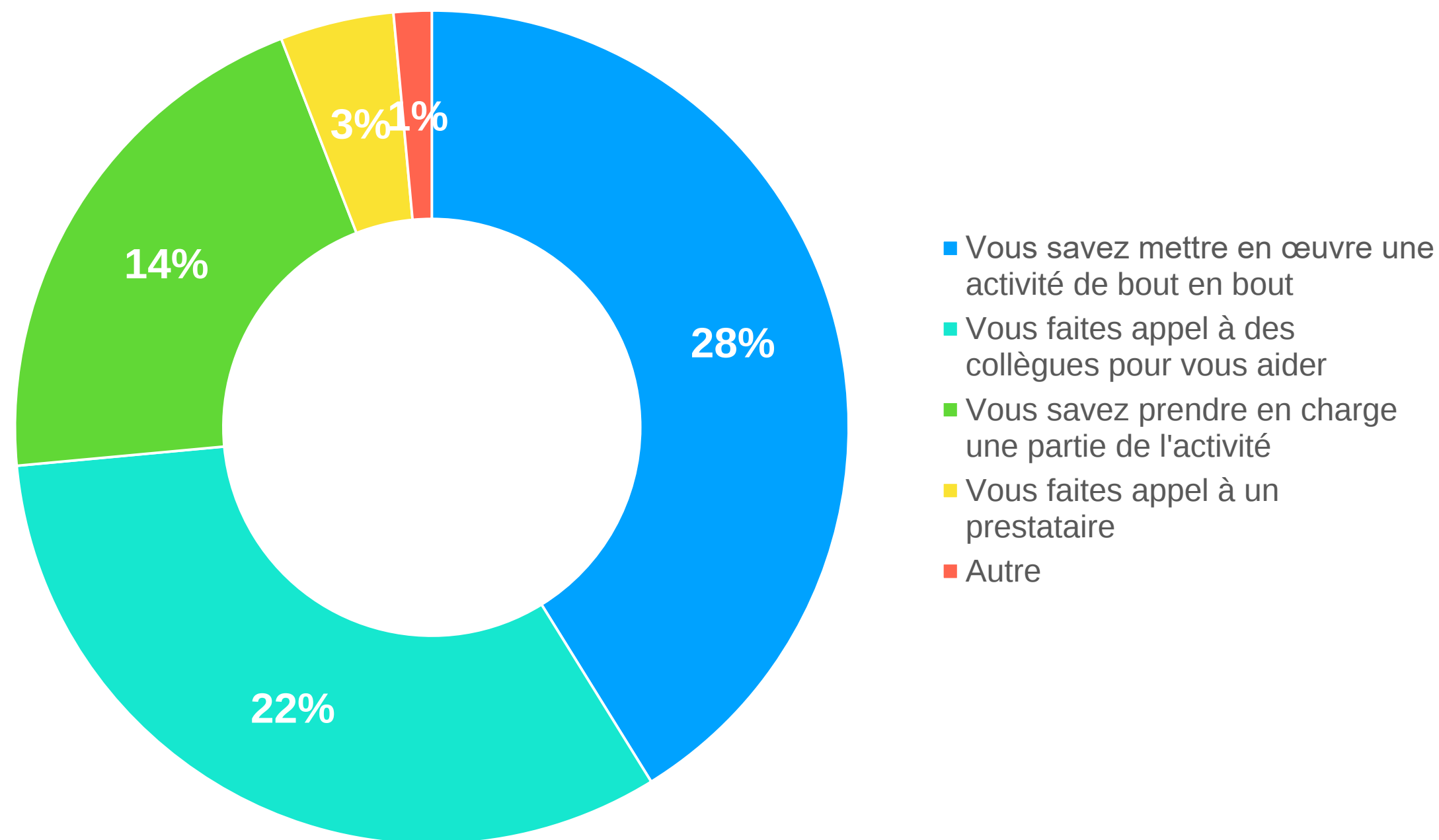
Estimez-vous connaître vos droits en matière de fouille de texte ? Profil des utilisateurs

Profil	Plutôt bien mais j'ai parfois des hésitations	J'en ai entendu parler	Pas du tout
Chercheur	23%	26%	24%
Dirigeant	3,5%	1%	0%
Doctorant	5%	5%	10%
Enseignant	1%	3%	4%
Enseignant-chercheur	21%	11%	10%
Etudiant	1%	1%	2%
Fonction d'Appui	24%	20,5%	26%
Technicien	5%	4%	2%
Autre	7%	11%	10%

235 réponses



Pour vos activités de fouille de textes en général :



28%

Vous savez mettre en œuvre une activité de bout en bout

22%

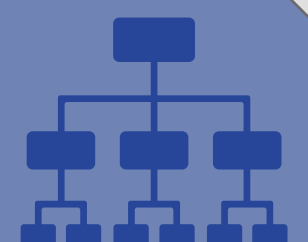
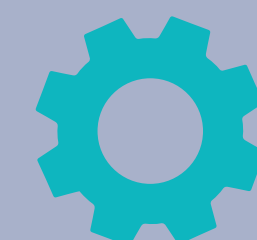
Vous faites appel à des collègues pour vous aider

14%

Vous savez prendre en charge une partie de l'activité

AUTRE

Se limiter faute de formation, navigation à vue, solution installée chez les partenaires

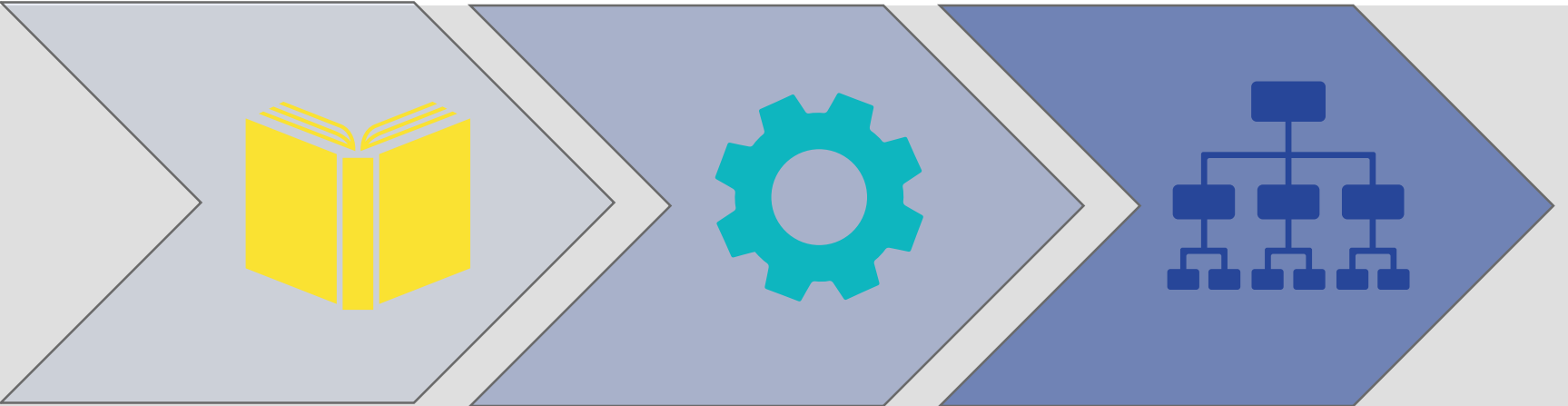


Pour vos activités de fouille de textes en général / profil des utilisateurs :

Profil	Vous savez mettre en œuvre une activité de bout en bout	Vous faites appel à des collègues pour vous aider	Vous savez prendre en charge une partie de l'activité
Chercheur	30%	30%	29,5%
Dirigeant	5%	4%	4,5%
Doctorant	13%	4%	11%
Enseignant	3,5%	3%	2%
Enseignant-chercheur	20%	18%	23%
Etudiant	1%	4%	0%
Fonction d'Appui	9,5%	16%	18%
Technicien	6%	7%	7%
Autre	8%	9%	7%

AUTRE

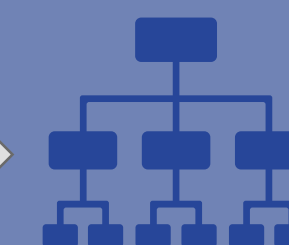
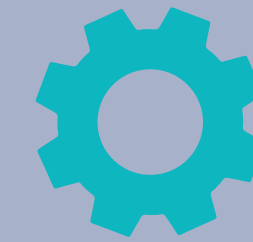
Se limiter faute de formation, navigation à vue, solution installée chez les partenaires



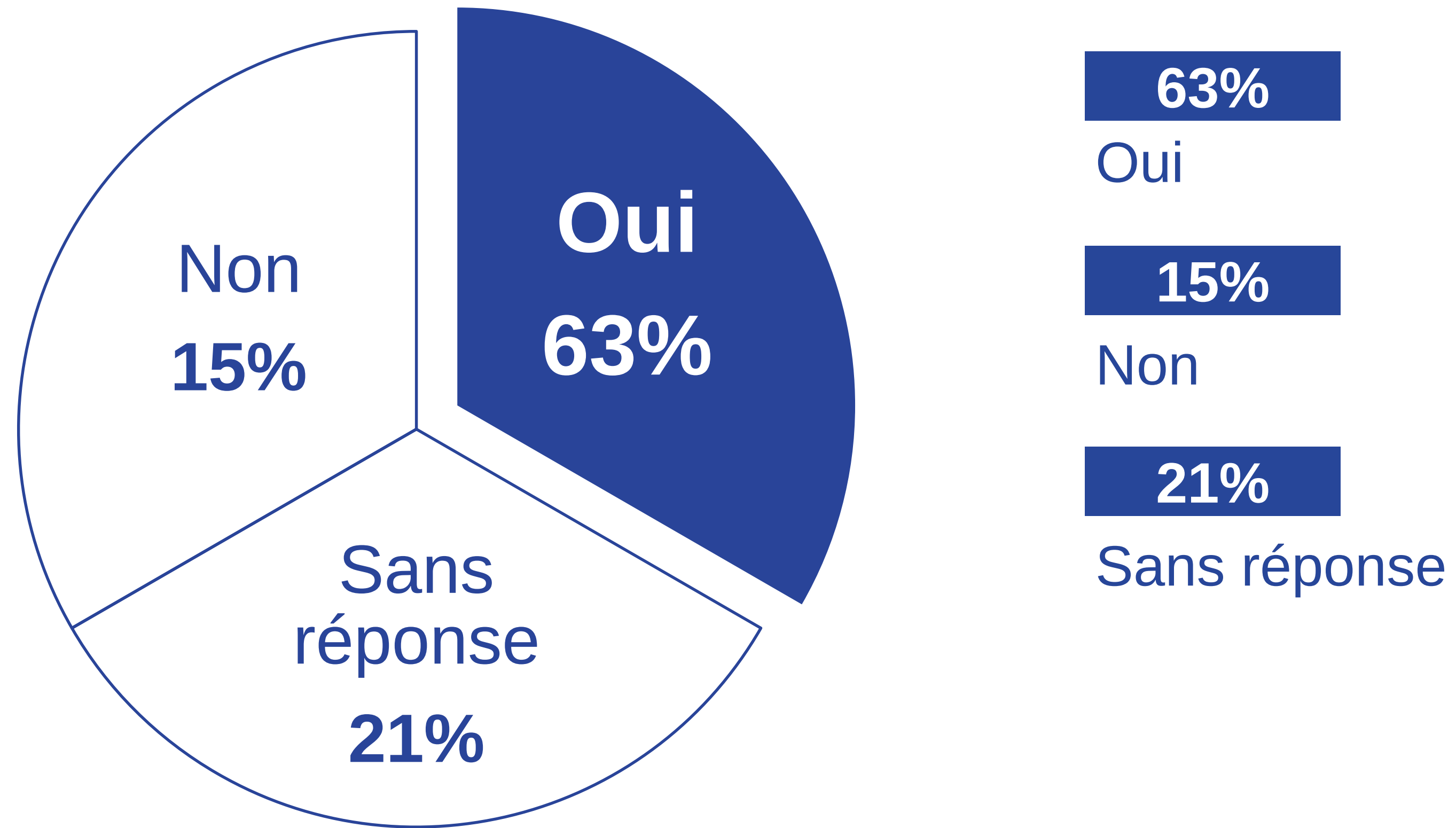
Pour vos activités de fouille de textes en général / niveau de formation :

Niveau de formation	Vous savez mettre en œuvre une activité de bout en bout	Vous faites appel à des collègues pour vous aider	Vous savez prendre en charge une partie de l'activité	Vous faites appel à un prestataire
Auto-formation	48%	36%	31%	5%
Diplôme Universitaire	61%	34%	16%	2%
Simple information	4%	20%	9%	7%
Formation(s) ponctuelle(s)	43,5%	49%	38%	2,5%
Aucune connaissance en fouille de textes	3%	8,5%	3%	0%

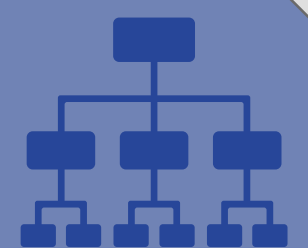
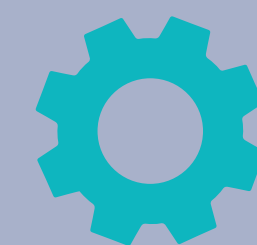
Croisement des données entre le niveau de formation et leur aisance dans les activités de fouille de textes



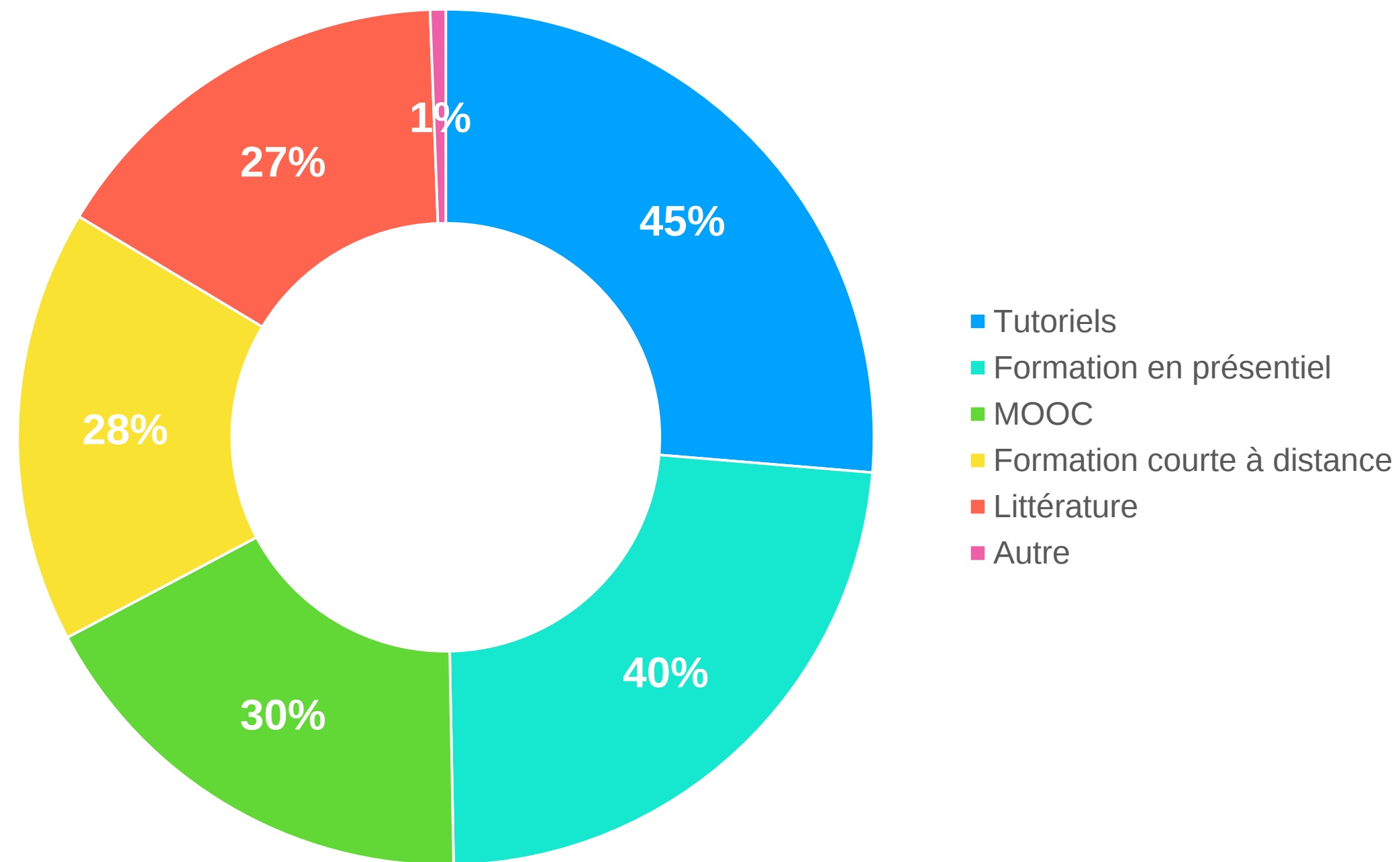
Pourriez-vous envisager de vous former pour développer vos compétences en fouille de textes ?



234 réponses



Quelles sont les modalités de formation qui vous semblent les plus adaptées à votre activité pour développer les compétences nécessaires à la mise en œuvre de projets de fouille de textes ?



45%

Tutoriels

40%

Formation en présentiel

30%

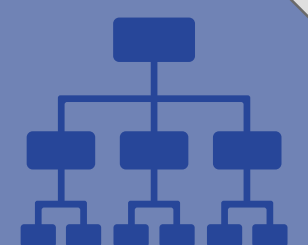
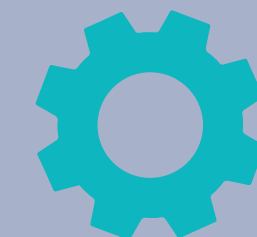
MOOC

28%

Formation courte à distance

AUTRE

Ateliers, groupe d'échanges, démos, co-développement professionnel

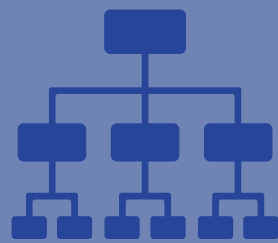


Quelles sont les modalités de formation qui vous semblent les plus adaptées à votre activité pour développer les compétences nécessaires à la mise en œuvre de projets de fouille de textes ? / profil des utilisateurs

Profil	Tutoriels	Formation en présentiel	MOOC	Formation courte à distance	Littérature
Chercheur	19%	21%	20%	21%	28%
Dirigeant	2%	2%	2%	3%	2%
Doctorant	10%	10%	13%	13%	12%
Enseignant	3%	2%	2%	2%	4%
Enseignant-chercheur	16%	14%	14%	14%	19%
Etudiant	1%	2%	1%	0%	0%
Fonction d'Appui	25%	25%	24%	27%	20%
Technicien	6%	5%	5%	6%	7%
Autre	8%	11,5%	10%	12%	4%

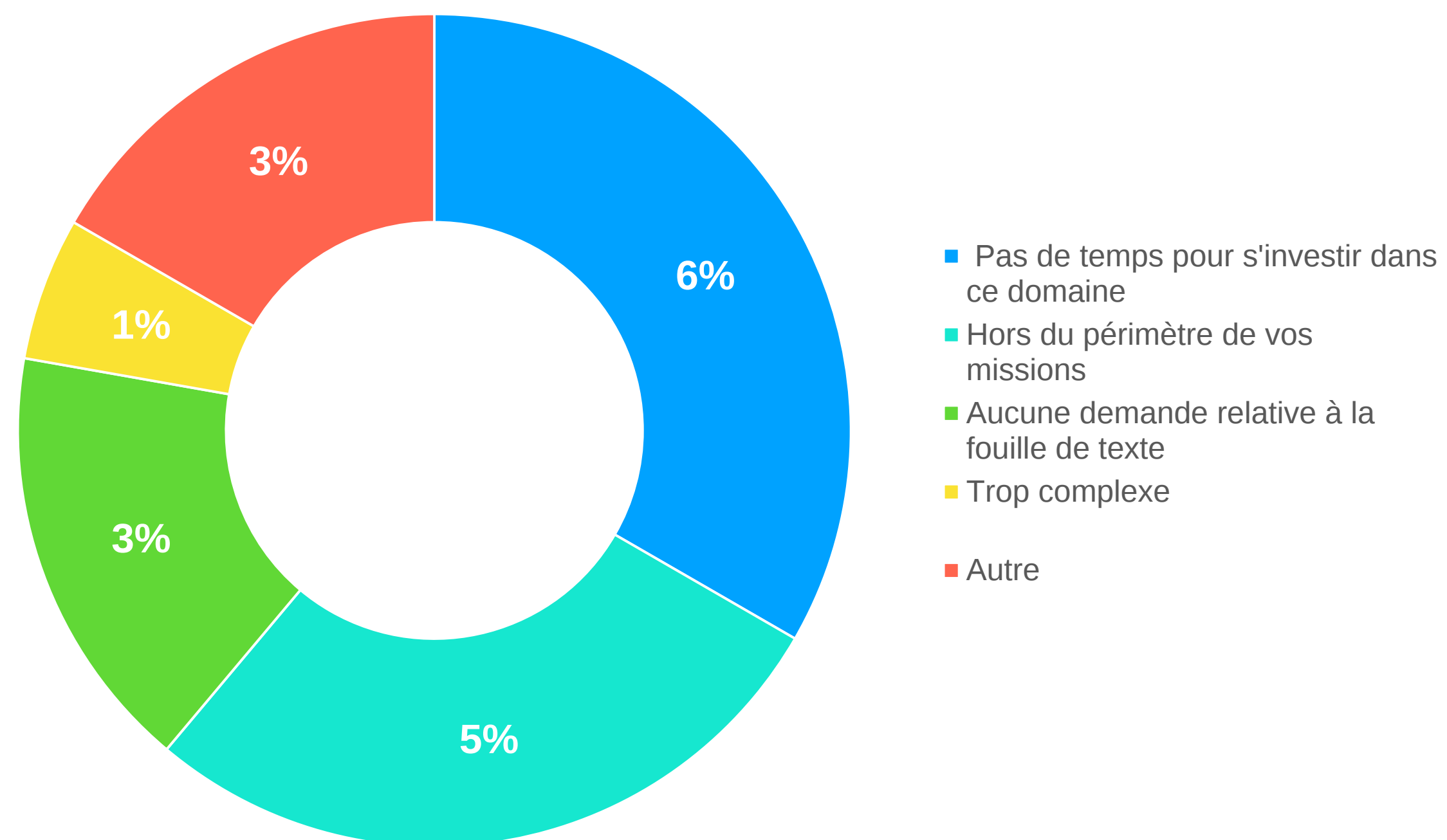
AUTRE

Ateliers, groupe d'échanges, démos, co-développement professionnel



Pour quelle(s) raison(s) ?

60
réponses



6%

Pas de temps pour s'investir dans ce domaine

5%

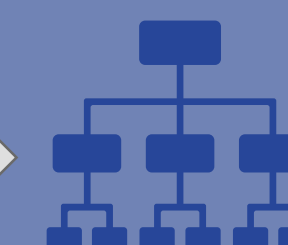
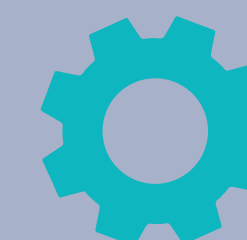
Hors du périmètre de vos missions

3%

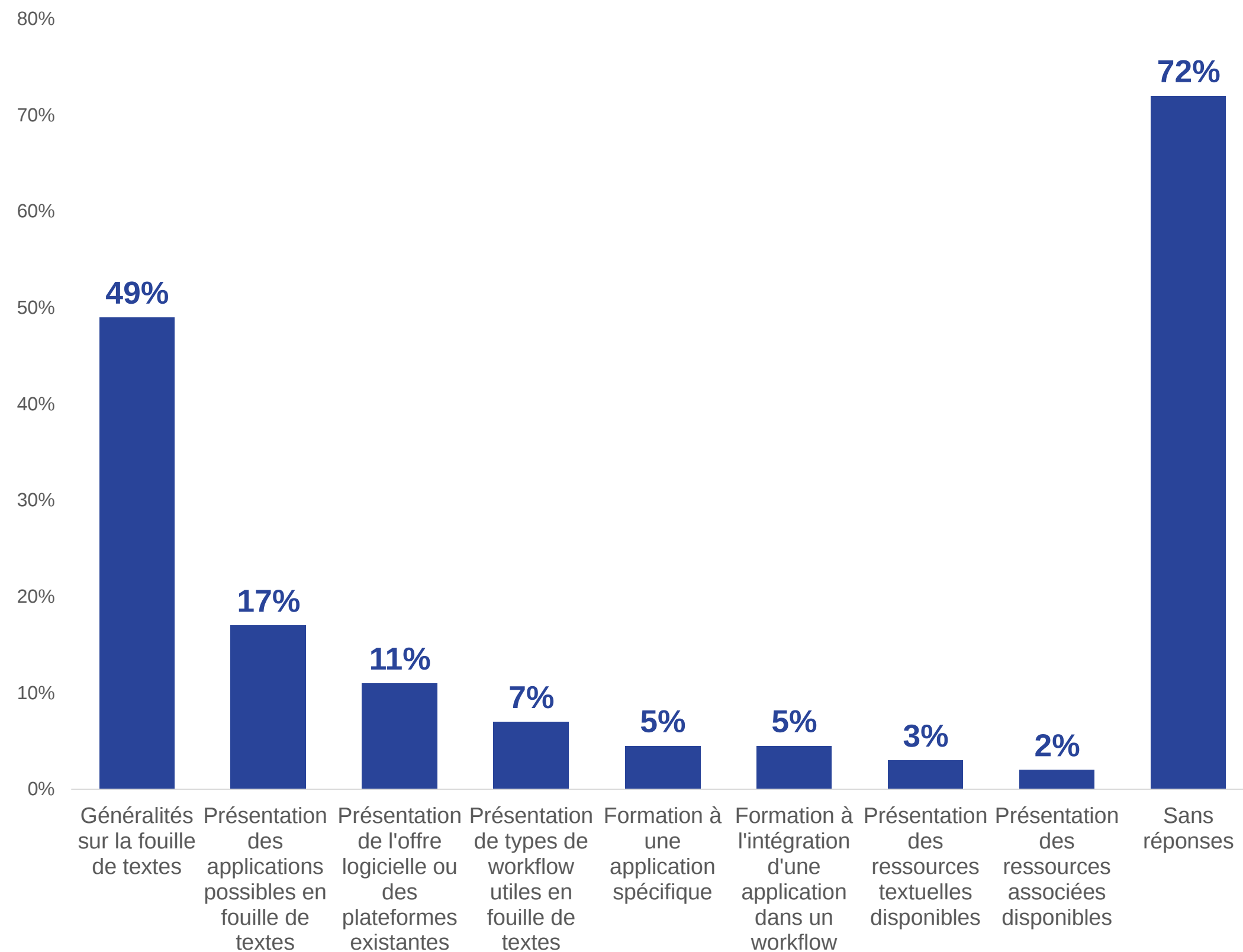
Aucune demande relative à la fouille de textes

AUTRE

Autoformation, pratique, pas de corpus à exploiter



Quels types de contenus devraient être abordés selon vous par ces formations ?



49%

Généralités sur la fouille de textes

17%

Présentation des application possibles en fouilles de texte

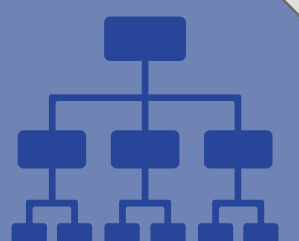
11%

Présentation de l'offre logicielle ou des plateformes existantes

7%

Présentation de types de workflow utiles en fouilles de textes

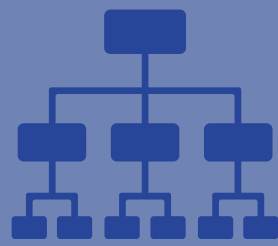
72% sans réponse



Quels types de contenus devraient être abordés selon vous par ces formations ? / profil des utilisateurs

Profil	Généralités sur la fouille de textes	Présentation des applications possibles en fouille de textes	Présentation de l'offre logicielle ou des plateformes existantes	Présentation des types de workflows utiles en fouille de textes	Formation à une application spécifique	Formation à l'intégration d'une application dans un workflow	Présentation des ressources textuelles disponibles	Présentation des ressources associées disponibles
Chercheur	20%	13,5%	19%	25%	35%	13%	19%	15%
Dirigeant	2%	5%	0%	0%	0%	6%	3%	2,5%
Doctorant	5%	5%		8%	4%	3%	3%	8%
Enseignant	3,5%	0%	7%	0%	4%	3%	0%	0%
Enseignant-chercheur	15%	13,5%	12%	4%	11,5%	19%	3%	10%
Etudiant	0%	0%	0%	0%	0%	0%	3%	0%
Fonction d'Appui	25%	22%	26%	21%	23%	22,5%	22%	18%
Technicien	3,5%	0%	2%	0%	4%	0%	6%	2,5%
Autre	12%	13,5%	7%	25%	8%	13%	6%	1,5%

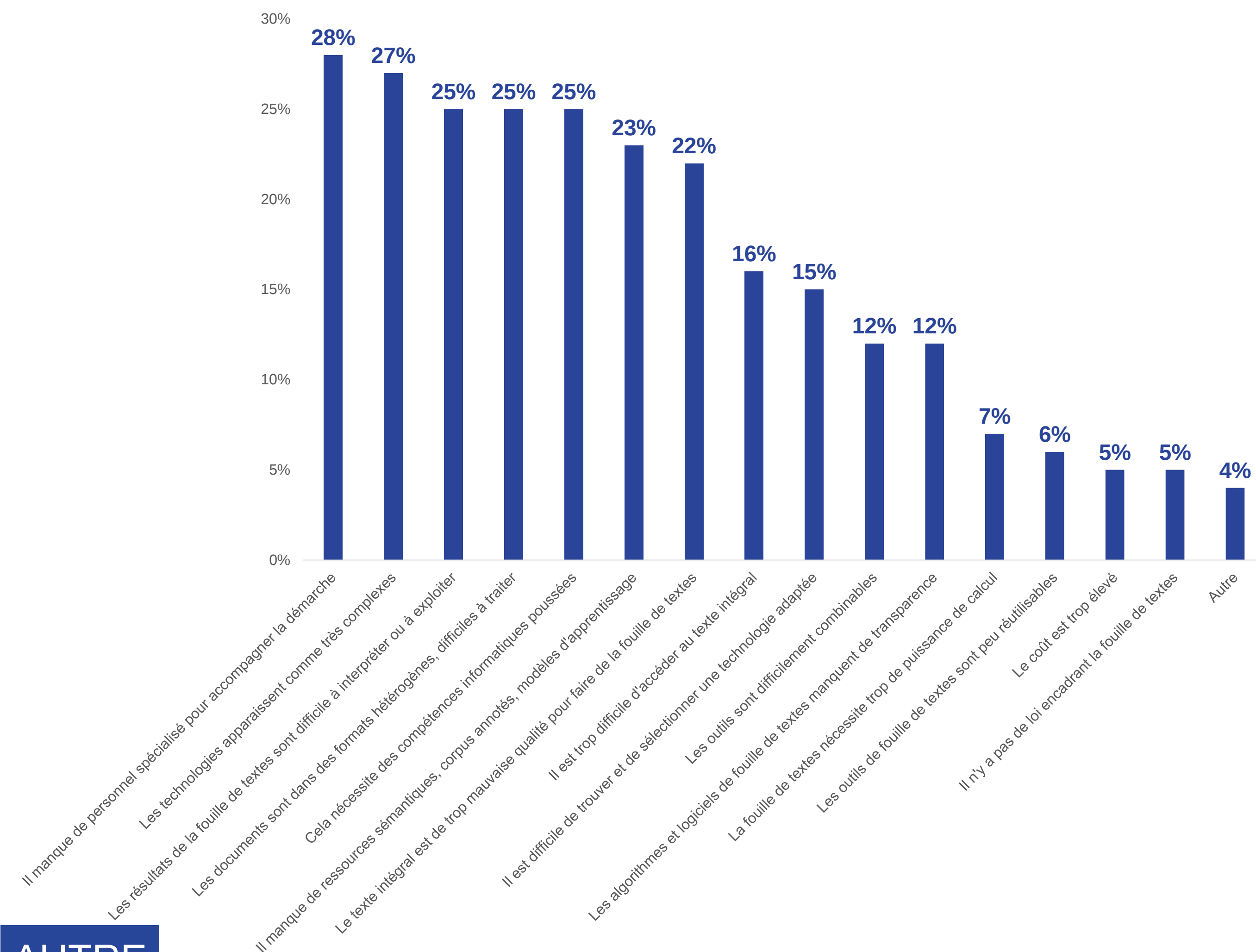
72% sans réponse



Freins et leviers

Quels sont, selon vous, les principaux freins actuels à la fouille de textes ?

300
réponses



28%

Manque de personnel spécialisé pour accompagner la démarche

27%

Les technologies apparaissent comme très complexes

25%

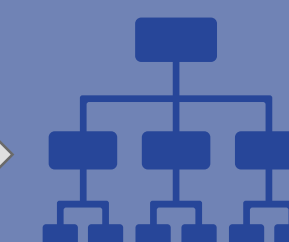
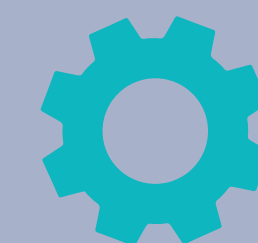
Documents dans des formats hétérogènes difficile à traiter

25%

Cela nécessite des connaissances informatique poussées

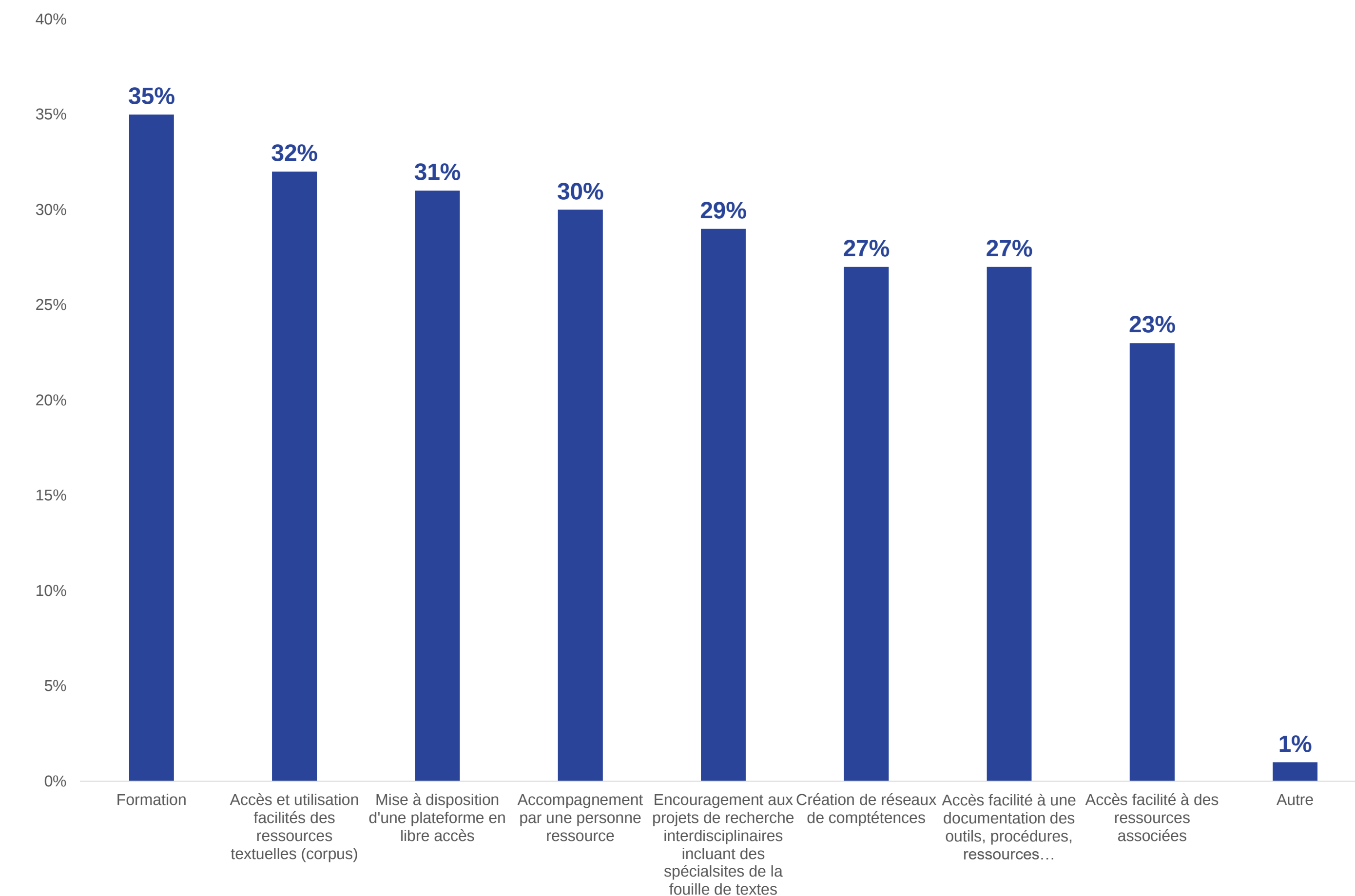
AUTRE

Manque de ressource en Français, absence de guide méthodologique, difficile d'imaginer la potentialité de la fouille de textes



Quelles principales mesures pourraient être utiles pour lever ces freins ?

300
réponses



35%

La formation

32%

Accès et utilisation facilités des ressources textuelles

31%

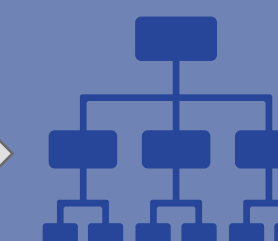
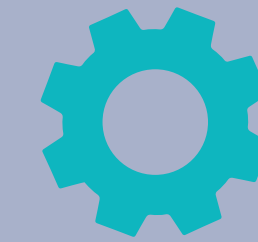
Mise à disposition d'une plateforme en libre accès

30%

Accompagnement par une personne ressource

AUTRE

Clarification sur la légalité, recrutement



Fonctionnalités d'une infrastructure

Comment verriez-vous les fonctionnalités d'une infrastructure de fouille de textes ?

1/2

OPERATIONS DE FOUILLE DE TEXTES

29%

Identifier la langue d'un texte
Indispensable

34%

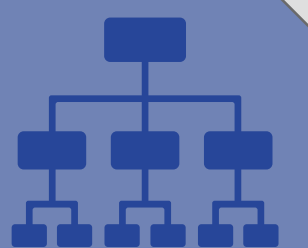
Annoter / extraire des entités nommées
Indispensable

34%

Annoter / extraire et standardiser des termes, des mots-clés
Indispensable

Choix d'une réponse entre Indispensable / Utile / Pas utile

14% sans réponse



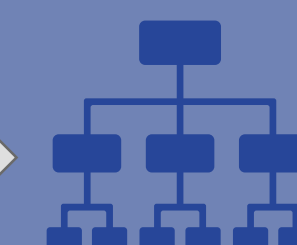
Comment verriez-vous les fonctionnalités d'une infrastructure de fouille de textes ?

2/2

OPERATIONS DE FOUILLE DE TEXTES

- 29% Produire une analyse morphologique, syntaxique
Utile
- 29% Classer, catégoriser des documents, identifier la thématique
Utile
- 30% Annoter / extraire des évènements
Utile
- 28% Analyser le discours, les sentiments
Utile

14% sans réponse



Comment verriez-vous les fonctionnalités d'une infrastructure de fouille de textes ?

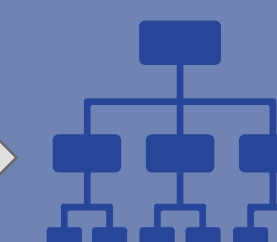
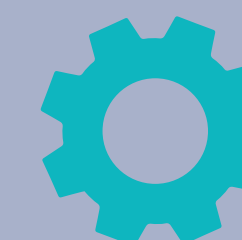
1/3

FONCTIONS DE BASE

- 19%** Configurer et adapter des processus de traitements existants
Indispensable
- 18%** Utiliser un processus existant avec un paramétrage défini
Utile
- 17%** Connaître l'état d'avancement d'un traitement que j'ai lancé
Utile

Choix d'une réponse entre Indispensable / Utile / Pas utile

32% sans réponse



Comment verriez-vous les fonctionnalités d'une infrastructure de fouille de textes ?

2/3

FONCTIONS DE BASE

29%

Exporter des résultats de traitement dans des formats standards
Indispensable

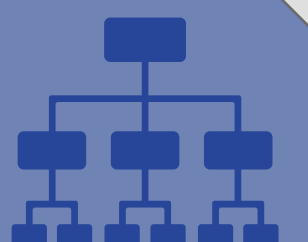
18%

Pouvoir interrompre un traitement et le reprendre
Utile

18%

Mixer une annotation manuelle avec une annotation automatique
Utile

32% sans réponse



Comment verriez-vous les fonctionnalités d'une infrastructure de fouille de textes ?

3/3

FONCTIONS DE BASE

17%

Exporter des résultats intermédiaires du processus de traitement

Utile

20%

Disposer d'une importante puissance de calcul

Utile

19%

Pouvoir traiter rapidement de gros volumes

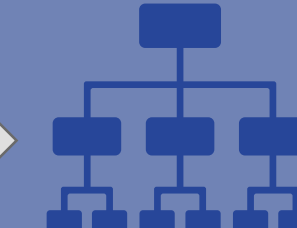
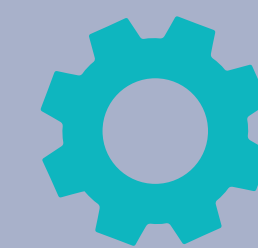
Utile

19%

Disposer d'un important espace de stockage

Utile

32% sans réponse



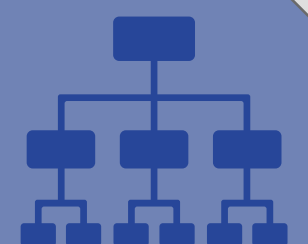
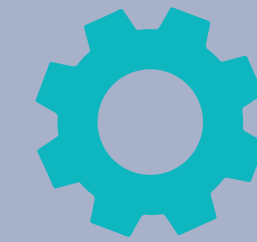
Comment verriez-vous les fonctionnalités d'une infrastructure de fouille de textes ?

INTEGRATION

- 9% Disposer de fonctionnalités de base à travers une API documentée
Indispensable
- 12% Intégrer facilement de nouveaux outils
Utile
- 9% Accéder à un environnement de test lors du dépôt d'outils
Utile

Choix d'une réponse entre Indispensable / Utile / Pas utile

50% sans réponse



Comment verriez-vous les fonctionnalités d'une infrastructure de fouille de textes ?

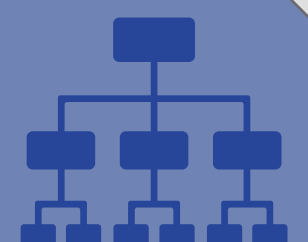
1/2

CORPUS

- 26%** Traiter un corpus composé de documents dans des formats hétérogènes
Indispensable
- 39%** Charger son propre corpus
Indispensable
- 25%** Constituer et traiter un corpus à partir de plusieurs sources connectées à la plateforme
Utile

Choix d'une réponse entre Indispensable / Utile / Pas utile

17% sans réponse



Comment verriez-vous les fonctionnalités d'une infrastructure de fouille de textes ?

2/2

CORPUS

CORPUS

30%

Disposer d'outils pour préparer un corpus (mise au format, découpage, ...)

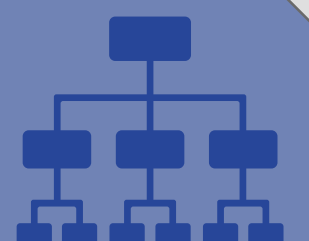
Indispensable

27%

Mettre à jour automatiquement des corpus disponibles sur la plateforme

Utile

17% sans réponse



Comment verriez-vous les fonctionnalités d'une infrastructure de fouille de textes ?

RESSOURCES ASSOCIEES

18%

Adapter des traitements à l'aide d'une ressource langagière, sémantique, statistique

Utile

22%

Accéder directement à des ressources utiles pour adapter les traitements

Utile

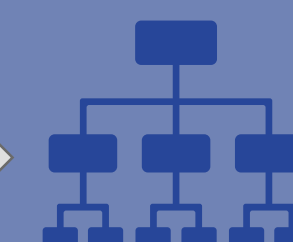
19%

Disposer d'outils et d'interfaces pour travailler les ressources directement sur la plateforme

Utile

Choix d'une réponse entre Indispensable / Utile / Pas utile

33% sans réponse



Comment verriez-vous les fonctionnalités d'une infrastructure de fouille de textes ?

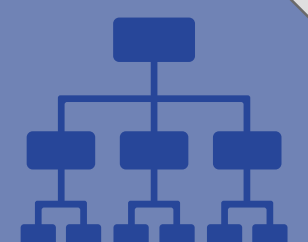
1/3

COLLABORATIF

- 27% Disposer d'un espace personnel de travail sécurisé
Indispensable
- 32% Partager des outils, ressources ou des traitements que je développe
Utile
- 29% Partager un corpus (entrée) avec d'autres utilisateurs
Utile

Choix d'une réponse entre Indispensable / Utile / Pas utile

18% sans réponse



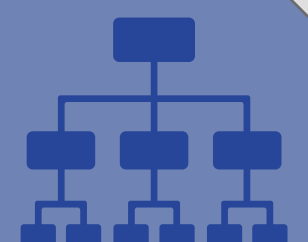
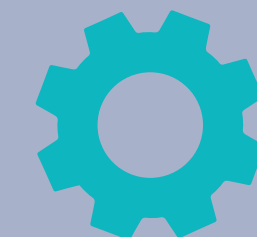
Comment verriez-vous les fonctionnalités d'une infrastructure de fouille de textes ?

2/3

COLLABORATIF

- 28%** Partager les résultats de mes traitements avec d'autres utilisateurs
Utile
- 27%** Accéder à des outils développés par d'autres
Utile
- 26%** Être informé de l'utilisation de mes ressources / outils
Utile

18% sans réponse



Comment verriez-vous les fonctionnalités d'une infrastructure de fouille de textes ?

3/3

COLLABORATIF

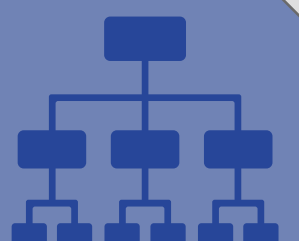
27%

Créer des groupes d'utilisateurs identifiés
Utile

32%

Disposer d'un forum de discussion
Utile

18% sans réponse



Comment verriez-vous les fonctionnalités d'une infrastructure de fouille de textes ?

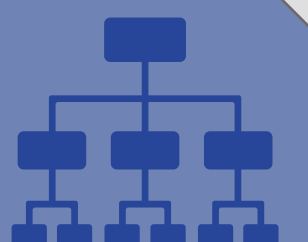
1/2

INTERFACES

- 27% Visualiser les résultats sous forme brute
Indispensable
- 24% Accéder depuis une même interface aux outils TM, aux corpus et aux ressources
Utile
- 28% Composer de manière graphique un processus de traitements
Utile

Choix d'une réponse entre Indispensable / Utile / Pas utile

19% sans réponse



Comment verriez-vous les fonctionnalités d'une infrastructure de fouille de textes ?

2/2

INTERFACES

28%

Visualiser les résultats sous forme graphique, dynamique

Utile

27%

Visualiser les résultats intermédiaires du processus de traitement

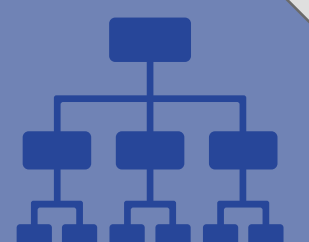
Utile

27%

Disposer d'interfaces adaptées en fonction de mon profil utilisateur

Utile

19% sans réponse



Comment verriez-vous les fonctionnalités d'une infrastructure de fouille de textes ?

AIDE

29%

Accéder à une documentation claire et complète, adaptées à mon profil

Indispensable

40%

Accéder à ne documentation claire et complète des outils et ressources disponibles

Indispensable

29%

Avoir une Foire Aux Questions (FAQ)

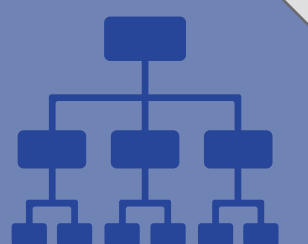
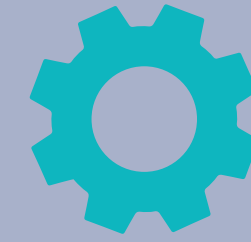
Utile

28%

Avoir un guichet support

Utile

16% sans réponse



Faites-nous part de vos recommandations ou autres attentes vis-à-vis des fonctionnalités d'une e-infrastructure pour la fouille de textes?

1/2

« Assurer la protection et la pérennité des données (accès, utilisation) chargées sur la plateforme. Assurer la mise à jour des outils proposés. Définir en amont une politique claire des conditions d'accès et d'utilisation de la plateforme (éligibilité des utilisateurs, coûts d'utilisation, limites d'espaces et techniques, choix d'évolution, etc.) »

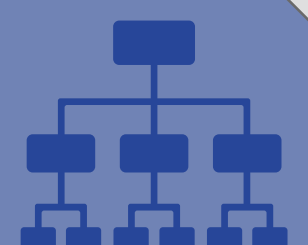
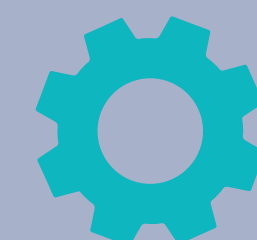
« Doit aussi se penser par rapport aux fonctionnalités existantes (qu'est ce qu'on peut apporter de plus que HyperbaseWeb, Cortext, AnalyseSHS ...) sinon risque que chacun ré-invente la roue dans son coin. »

La plupart des items sont sans réponse en raison du profil du répondant (documentaliste novice) qui exprime un besoin en formation pour accompagner des utilisateurs potentiels du TM et qui n'a pas besoin de TM pour ses besoins propres.

« Si possible couplée à logiciel de retranscription des entretiens. »

8 réponses

96% sans réponses



Faites-nous part de vos recommandations ou autres attentes vis-à-vis des fonctionnalités d'une e-infrastructure pour la fouille de textes?

2/2

« Vos questions sont déjà trop orientées vers votre plateforme, et non plus vers la fouille de texte en générale. Et ces questions sont beaucoup trop hors contexte pour que les réponses ne soient réellement utiles. Vous ne pouvez pas envisager une plateforme de fouille de textes qui fait tout ce que veut un utilisateur, chaque utilisateur ayant ses propres demandes, chaque corpus ayant ses propres spécificités. »

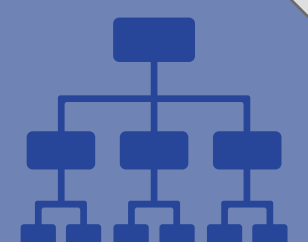
« Briser l'hégémonie du BigData a-sémantique, y compris en analyse d'opinion. »

« Le textmining va se généraliser à grande vitesse pour que tout un chacun puisse enfin exploiter facilement tous types de contenus textuels. À l'heure de l'infobesité les outils doivent être à la portée d'utilisateurs comme le sont les nouvelles générations d'application. »

« Pas spécialiste en analyse des données textuelles (text mining) j'aimerais avoir à ma disposition un outil aussi facile d'emploi qu'un moteur de recherche, mais qui gère non pas seulement des mots-clés mais des concepts (ex : concept de résilience) ou des réseaux de concepts qui se forment et se déforment au cours du temps, avec des relations d'étymologie, de proximité, d'inclusion, et peut-être d'autres auxquelles je n'ai pas pensé. »

8 réponses

96% sans réponses

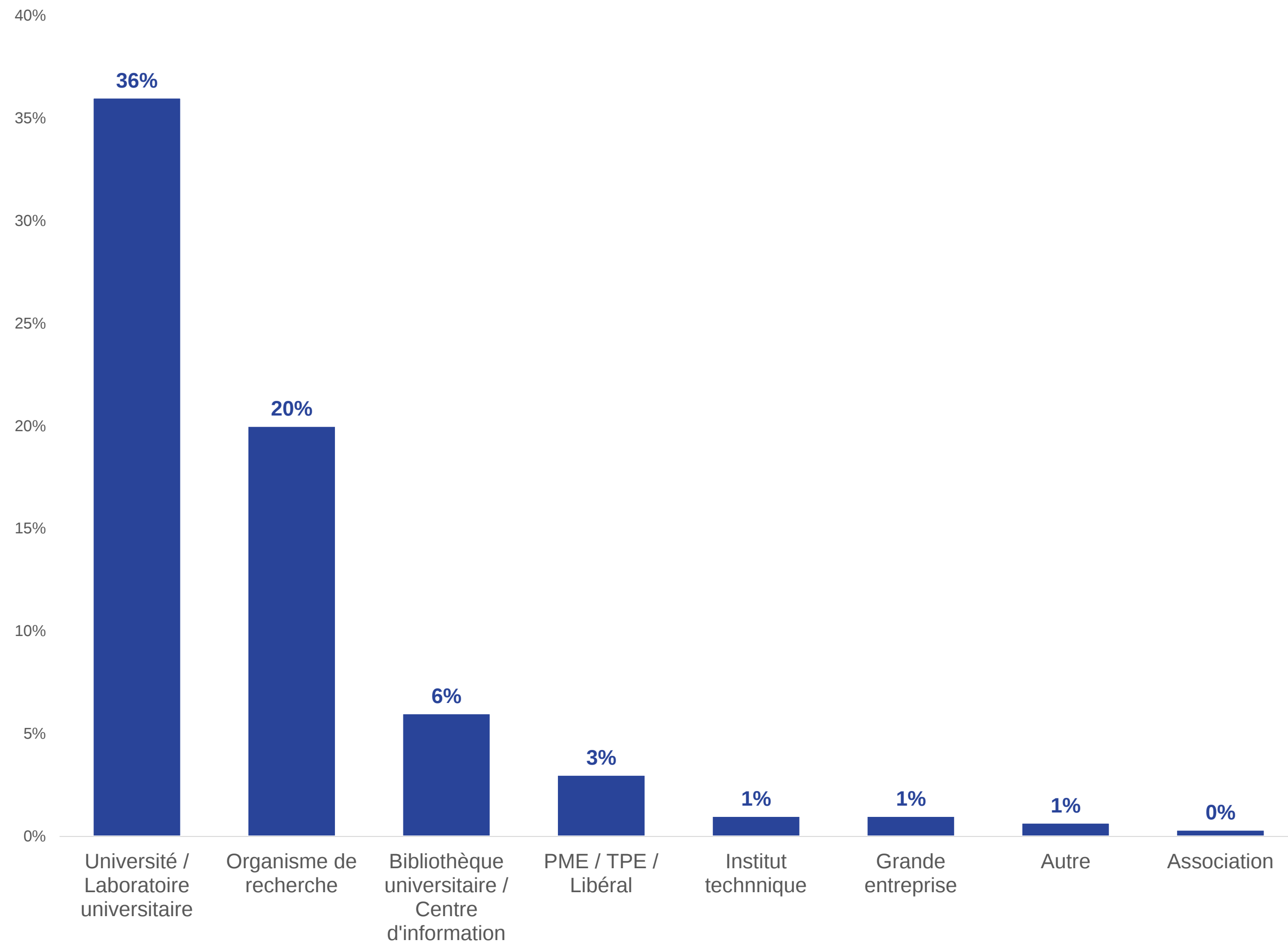


Profil

Des utilisateurs

Dans quel type d'organisation exercez-vous ?

209
réponses



36%

Université / Laboratoire universitaire

20%

Organisme de recherche

6%

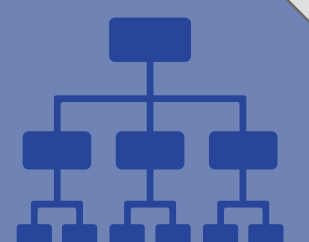
Bibliothèque universitaire / Centre d'information

3%

PME / TPE / Libéral

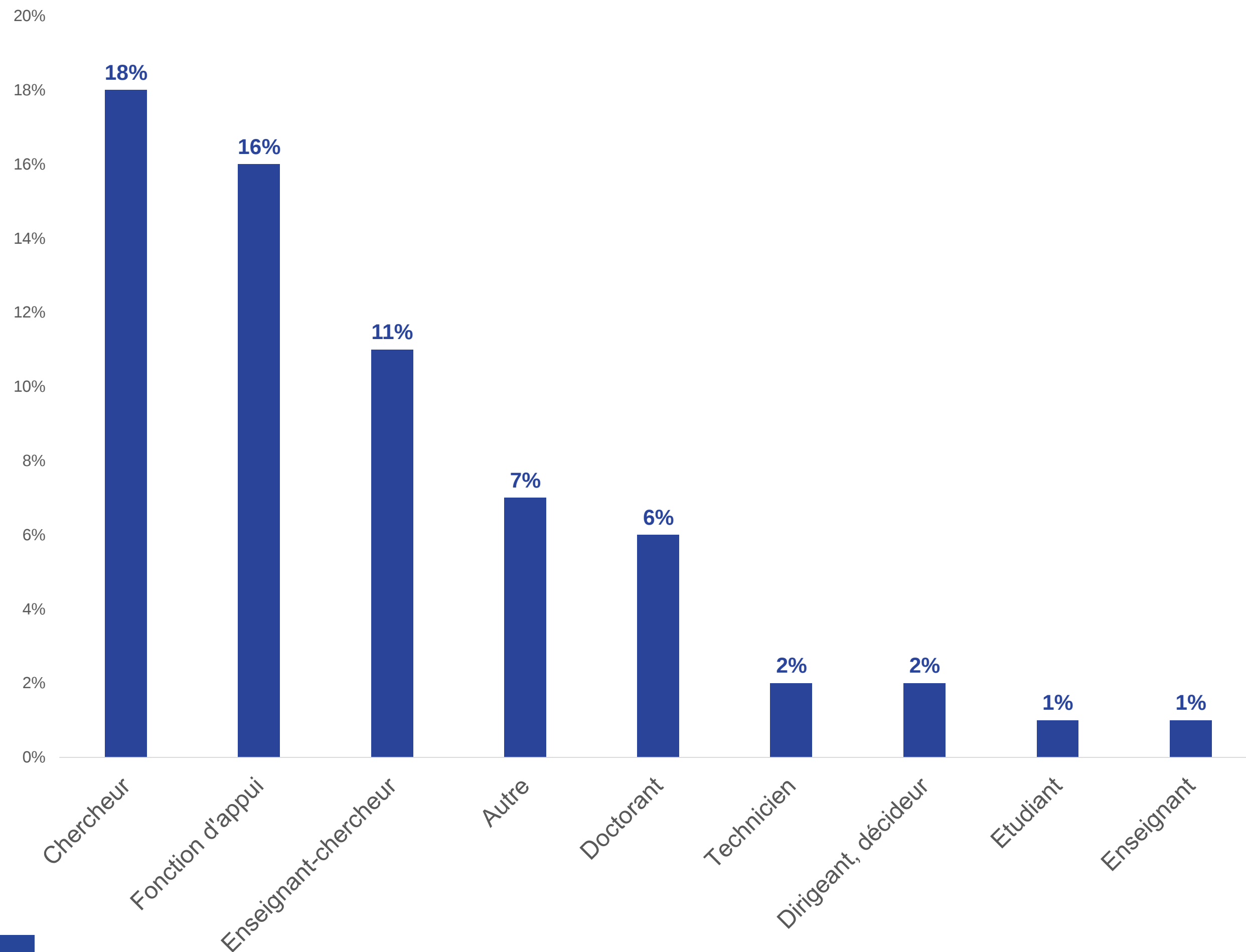
AUTRE

éducation nationale, startup



Quelle fonction / poste occupez-vous ?

202
réponses



18%

Chercheur

16%

Fonction d'appui

11%

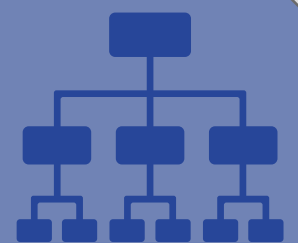
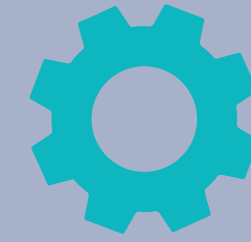
Enseignant / Chercheur

6%

Doctorant

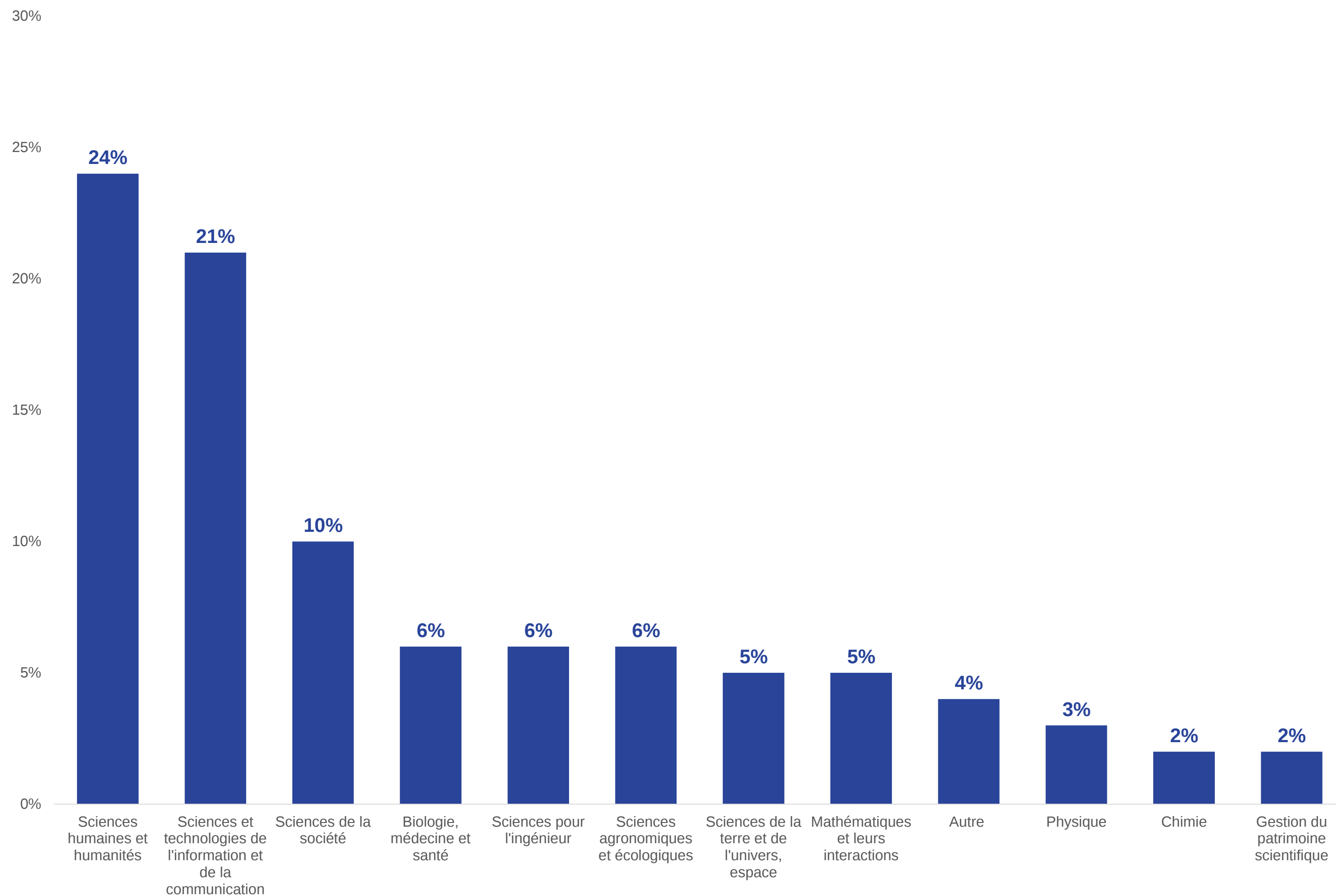
AUTRE

Chef de projet, ingénieur d'étude, ingénieur de recherche, ingénieur data, documentaliste, consultant département R&D, chargé de projet dans les outils numériques, architecte, développeur



Domaine(s) scientifique(s) ?

295
réponses



24%

Sciences Humaines et Humanités

21%

Sciences et technologies de l'information et de la communication

10%

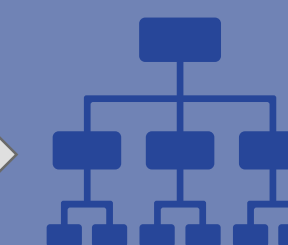
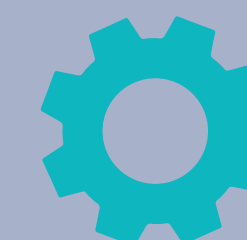
Sciences de la société

6%

Sciences agronomiques et écologiques

AUTRE

IST, informatique, droit, traitement automatique des langues, linguistique, sciences juridiques, ingénierie des connaissances, pluridisciplinaire



Persona

Des utilisateurs

Compétences et formation

- **Niveau de formation** : simple information et n'a aucune connaissance en fouille de textes
- Il souhaite se former pour développer ses compétences
- Les tutoriels et les formations en présentiel sont les modalités les plus adaptées à son profil
- Les généralités sur la fouille de texte, la présentation de l'offre logicielle ou des plateformes existantes et la présentation des applications possibles sont les types de contenus qui devraient être abordés selon lui.

Les articles scientifiques, comptes rendus, rapports et les pages web constituent les supports de son activité.

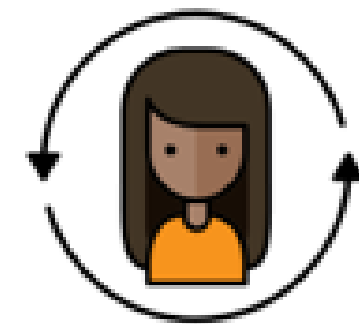
Fonctionnalités d'une infrastructure

Indispensable

- **Opération de fouille de textes** : Identifier la langue d'un texte, annoter / extraire des entités nommées, classer / catégoriser des documents
- **Corpus** : Constituer et traiter un corpus à partir de plusieurs sources, charger son propre corpus et disposer d'outils pour préparer un corpus
- **Collaboratif** : Disposer d'un espace personnel de travail sécurisé
- **Aide** : Accéder à une documentation claire, complète et adaptée à leur profil ainsi que l'accès à des outils et ressources disponibles

Utile

- Les éléments proposés pour **l'interface** d'une infrastructure sont considéré comme utile mais non indispensable



Non

3

UTILISATEUR

- Il exerce en université et en organisme de recherche
- Il occupe une fonction d'appui et de chercheur
- Il travaille en STIC et en Sciences Humaines et Humanités

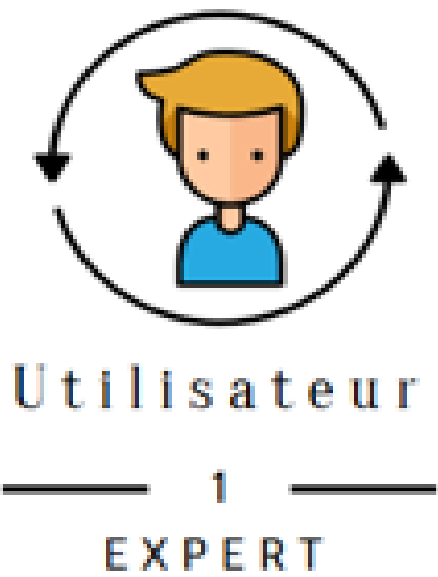
- Avoir des compétences informatiques poussées et la complexité des technologies représentent les principaux freins.
- L'accompagnement par une personne ressource et la formation sont des mesures utiles pour lever ces freins.

La fouille de texte dans vos activités

- Il utilise la fouille de textes pour l'extraction d'informations et la classification / catégorisation de documents
- Il utilise des solutions libres
- Il développe ses propres outils pour la fouille de textes
- Il applique la fouille de textes sur les articles scientifiques, les pages web et les réseaux sociaux
- Il applique des pré-traitements de type Pré-traitements linguistiques et transformation de format
- Les lexiques, listes de termes, modèles d'apprentissage et les corpus annotés constituent les ressources utilisées pour le text-mining
- Il développe ses propres ressources comme les corpus annotés et les modèles d'apprentissage
- Il a l'habitude de rendre ces productions accessibles et réutilisables par d'autres.

Fonctionnalités d'une infrastructure

Identifier la langue d'un texte, Annoter / extraire et standardiser des termes ou mots-clés et annoter / extraire des entités nommées sont les fonctionnalités indispensables des **opérations de fouille de textes**,
Les **fonctions de bases** indispensables sont :
- Utiliser un processus existant avec un paramétrage pré-défini
- Configurer et adapter des processus de traitement existants
- Connaître l'état d'avancement d'un traitement que j'ai lancé
- Exporter des résultats de traitement dans des formats standards
Disposer de fonctionnalités de base à travers une API documentée est indispensable pour **l'intégration** d'une infrastructure en fouille de textes.
Les éléments indispensables dans **un corpus** sont :
- Constituer et traiter un corpus à partir de plusieurs sources connectées à la plateforme
- Charger son propre corpus
- Disposer d'outils pour préparer un corpus
Adapter des traitements à l'aide d'une ressource langagière, sémantique, statistique s'avère indispensable pour les **ressources associées**.
Disposer d'un espace personnel de travail sécurisé est le seul élément indispensable pour un **espace collaboratif**. Les autres fonctionnalités sont décrites comme utile.
Visualiser les résultats sous forme brute s'avère indispensable pour les **interfaces** d'une infrastructure. Les autres éléments sont décrits comme utile.
Accéder à une documentation claire et complète adaptée au profil ainsi qu'à des outils et ressources disponibles sont perçu comme indispensable pour le module **aide** d'une infrastructure.



Utilisateur
1
EXPERT

Compétences et formation

- **Niveau de formation** : diplôme universitaire et auto-formation
- Il connaît ses droits en fouilles de textes mais a parfois des hésitations
- Il sait mettre en œuvre une activité de bout en bout
- Il veut se former pour développer ses compétences mais n'a pas de temps à investir dans ce domaine
- Les tutoriels et la littérature représentent les modalités les plus adaptées pour se former
- Les généralités sur la fouille de texte, la formation à une l'intégration d'une application dans un workflow et la présentation des types de workflow utiles en fouille de textes sont les types de contenus qui devraient être abordés selon lui.

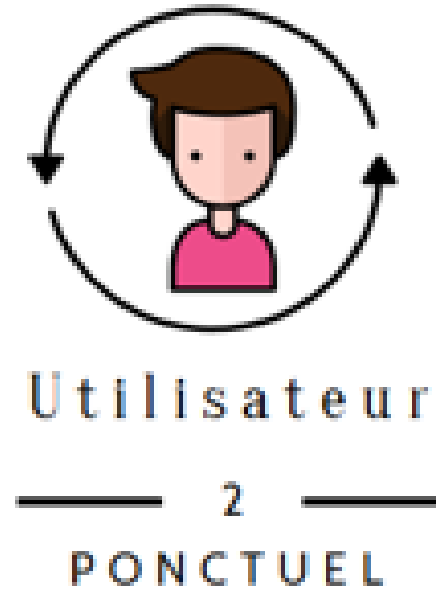
Freins et leviers

- Il exerce en université et en organisme de recherche
- Il est chercheur ou enseignant chercheur
- Il travaille en STIC et en Sciences Humaines et Humanités

- Les principaux freins actuels à la fouille de textes sont, selon lui :
Le manque de ressources sémantiques, corpus annotés et modèles d'apprentissage
La mauvaise qualité du texte intégral
- L'accès et l'utilisation facilitée de ressources textuelles ainsi que l'encouragement aux projets de recherches interdisciplinaires incluant des spécialistes de la fouille de textes sont des mesures utiles pour lever les freins.

La fouille de texte dans vos activités

- Il utilise la fouille de textes pour l'extraction d'informations, l'aide à la découverte scientifique et la classification / catégorisation de documents
- Il utilise des solutions libres
- Il applique la fouille de textes sur les articles scientifiques, les pages web, les réseaux sociaux et les comptes rendus / rapports
- Les lexiques, listes de termes, les corpus annotés et les Thésaurus, taxonomies constituent les ressources utilisées pour le text-mining
- Majoritairement, il ne développe pas ses propres ressources
- Il a l'habitude de rendre ces productions accessibles et réutilisables par d'autres



Les fonctionnalités indispensables **des opérations de fouille de textes** sont : **Fonctionnalités d'une infrastructure**

- Identifier la langue d'un texte
- Annoter / extraire et standardiser des termes ou mots-clés
- annoter / extraire des entités nommées
- Classer / catégoriser des documents et identifier des thématiques

Les **fonctions de bases** indispensables sont :

- Configurer et adapter des processus de traitement existants
- Exporter des résultats de traitement dans des formats standards

Le reste des éléments est désigné comme utile.

Les éléments indispensables dans **un corpus** sont :

- Traiter un corpus composé de documents dans des formats hétérogènes
- Charger son propre corpus
- Disposer d'outils pour préparer un corpus

Les ressources associées sont considérées comme utile mais non indispensable.

Disposer d'un espace personnel de travail sécurisé est le seul élément indispensable pour un **espace collaboratif**, les autres fonctionnalités sont décrites comme utile.

Accéder depuis une même interface aux outils TM, aux corpus et aux ressources ainsi que de visualiser les résultats sous forme brute s'avère indispensable pour les **interfaces** d'une infrastructure. Les autres éléments sont décrits comme utile.

Accéder à une documentation claire et complète adaptée au profil ainsi qu'à des outils et ressources disponibles sont perçus comme indispensables pour le module **aide** d'une infrastructure

Compétences et formation

- **Niveau de formation** : auto-formation et formation(s) ponctuelle(s)
- Il connaît ses droits en fouilles de textes mais a parfois des hésitations
- Il fait appel à des collègues pour être aidé dans ses activités de fouille de textes
- Les tutoriels et les formations en présentiel sont les modalités les plus adaptées pour se former
- Il veut se former mais celui-ci semble hors du périmètre de ses missions
- Les généralités sur la fouille de texte, la formation à une application spécifique et la présentation de l'offre logicielle ou des plateformes existantes sont les types de contenus qui devraient être abordés.

- Il exerce en université et en organisme de recherche
- Il est chercheur ou exerce une fonction d'appui
- Il travaille en STIC et en Sciences Humaines et Humanités

Freins et leviers

- Le manque de personnel spécialisé dans la démarche, la complexité des technologies et la difficulté à traiter des documents dans des formats hétérogènes sont les principaux freins
- La formation et la mise à disposition d'une plateforme en libre accès sont utiles pour lever les freins