



**CHAPSVISION**  
DATA MAKE SENSE

**SOLUTIONS IA ET SEMANTIQUES (QWAM NLP)**

**QWAM**  
by CHAPSVISION

## PLAN

- ❖ Introduction
  - ❖ Un ensemble de solutions et compétences NLP
  - ❖ Données textuelles, IA, sémantique
  
- ❖ QWAM Text Analytics (QWAM NLP)
  - ❖ Cas d'usage
  - ❖ Méthodologie
  - ❖ Démonstration API
  
- ❖ Solutions QWAM by Chapsvision
  - ❖ Module QWAM InfoMedia
  - ❖ Modules IA Générative : résumés et RAG\$
  - ❖ Illustrations : solutions intégrées avec Speech-to-text et traduction



#1

# DONNÉES TEXTUELLES, IA, SÉMANTIQUE...

## QUELLES DONNÉES TEXTUELLES DANS LES ORGANISATIONS ?



Fichiers et documents  
textes, contrats, rapports,...



Applications métiers et  
documentaires



Informations web &  
Réseaux sociaux



Emails



E-commerce, plateformes  
clients



Multimédia (transcriptions  
vidéo/audio, images)

**Ces données sont en général faiblement exploitées par manque de temps (traitement manuel) et d'outils. Elles sont simplement stockées, leur volume augmente et peu de valeur ajoutée ou d'indicateurs sont extraits.**

#2

# QWAM TEXT ANALYTICS

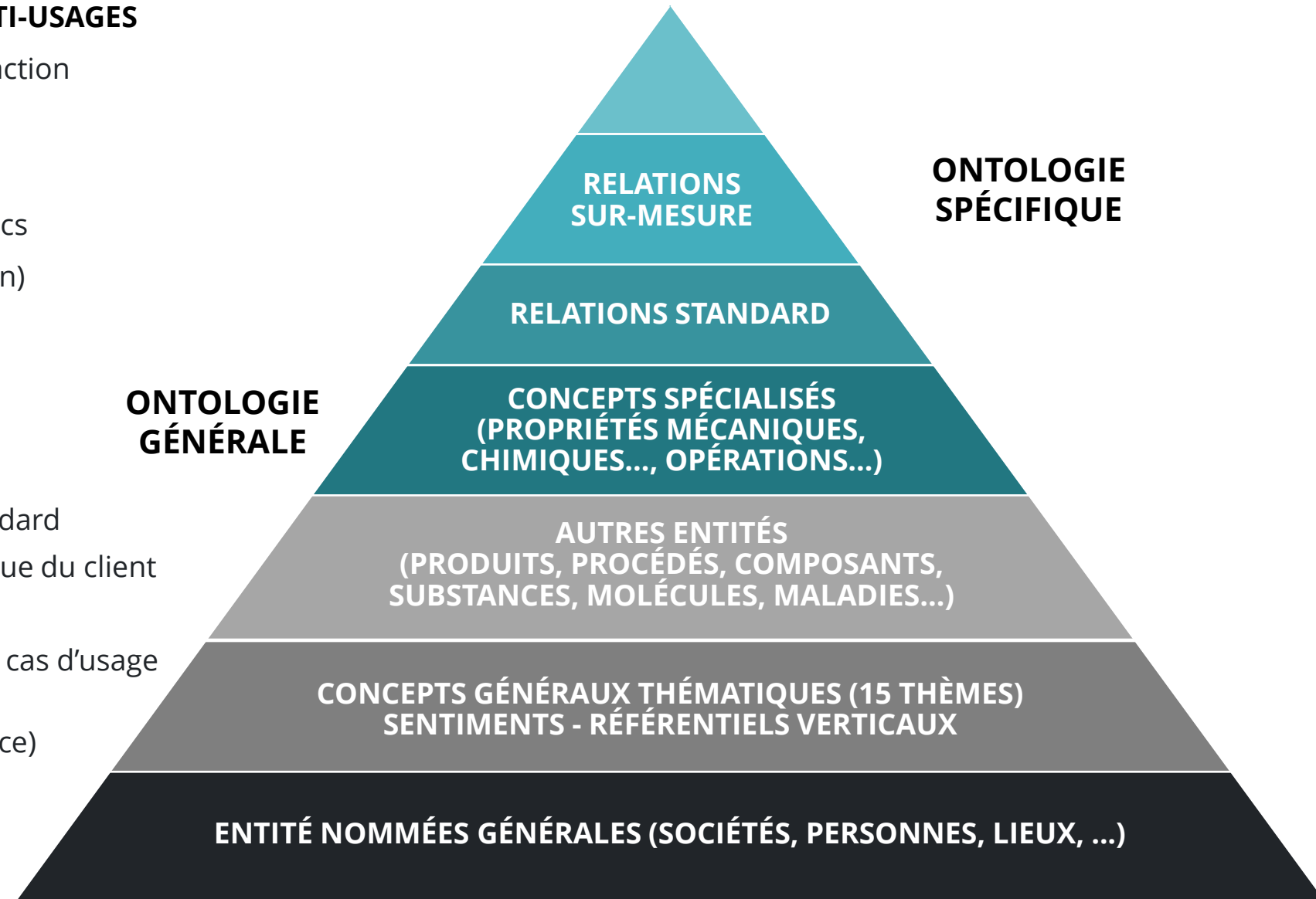
by CHAPSVISION

**SOLUTION HYBRIDE SÉMANTIQUE & IA MULTI-USAGES**

- Génération de métadonnées (tagging, extraction d'informations)
- Analyse de données textuelles
- Moteur de recherche sémantique et analytics
- Résumés et synthèses automatiques (IA Gen)
- RAG (Questions-réponses + IA Gen)

**AVANTAGES**

- Vocabulaire et modèle général livré en standard
- Adaptation très rapide au domaine spécifique du client
- Solution hybride sémantique/IA/Gen AI
- Vocabulaire et modèle spécialisé adapté au cas d'usage (contenus, indicateurs)
- Gestion d'ontologie et ressources (back-office)
- SaaS ou on-premise
- Plus de 150 applications réalisées



# EXTRACTION AUTOMATIQUE

- **Entités** nommées générales (sociétés, personnes, lieux, dates...)

+  
Autres entités (produits, procédés, composants, substances, molécules...)

+  
Concepts généraux thématiques (15 thèmes)

+  
Concepts spécialisés (propriétés mécaniques, chimiques..., opérations...)

+  
Relations standard

+  
Relations sur-mesure

+  
**Sentiments** généraux

+  
Sentiments spécifiques

- À partir de textes
- À partir de documents PDF, doc...

Concepts catégorisés / validés
Chiffre d'affaires (économie)
Livraison de colis (livraison)
Pandémie (santé)
Livraisons sans contact (transport)
E-commerce (économie)

Personnes
Boris Winkelmann ( sujet )
Philippe Wahl

Lieux
France
Allemagne
UK
Espagne
Europe
Pologne
Brésil
Afrique Du Sud
USA
Chine
Italie

Sociétés
La Poste ( sujet )
Chronopost
SEUR
La Banque Postale
Geopost ( sujet )
Deutsche Post
DHL
Fedex
UPS

GeoPost, la petite méconnue de La Poste qui pallie la chute du courrier

FOCUS-La filiale de livraison de colis, déjà présente dans 47 pays, veut doubler de taille d'ici 2025. Ses profits ont été multipliés par deux en 2020.

GeoPost dépassera sur l'année la barre des 10 milliards d'euros de chiffre d'affaires, avançant en juillet 2020 Boris Winkelmann, le PDG de cette filiale de La Poste. Ce spécialiste de la livraison de colis a fait mieux que tenir sa promesse: en 2020, ses revenus ont atteint 11 milliards grâce à une croissance de 42 % de l'activité.

«On a dû très vite s'adapter à la pandémie, mettre en place la livraison sans contact», raconte Boris Winkelmann. En mars, nos volumes ont baissé de 30 %. Mais, à partir de Pâques, il y a eu un rebond avec des performances historiques. Résultat, le roi du colis pour les entreprises et les particuliers pèse plus lourd (35 %) dans l'activité de La Poste que son métier historique, le courrier (18,7 %).

Pourtant, ce nom n'évoque rien pour le grand public. Et pour cause. Dans les 47 pays où elle déploie ses services, la société opère sous des marques locales: Chronopost, Chronofresh et Stuart en France, DPD en Allemagne et au Royaume-Uni, Seur en Espagne. Dans les prochaines années, cette société de 97.000 salariés devrait se faire un nom. Avec La Banque postale et les services numériques, la livraison de colis express doit servir d'amortisseur à la baisse de l'activité courrier pour La Poste d'ici 2030.

«Pour GeoPost, qui a livré 1,6 milliard de colis l'an passé, l'objectif est de passer numéro un en Europe devant Deutsche Post, DHL, expliquait il y a quelques jours au Figaro le PDG du groupe La Poste, Philippe Wahl. De fait, le roi de la livraison de colis à deux pas de chez soi ou à l'autre bout du monde compte accélérer ces prochaines années: il table sur un chiffre d'affaires de 21 milliards en 2025, soit presque un doublement par rapport à 2020. En 2030, il voudrait atteindre les 27 milliards de revenus.

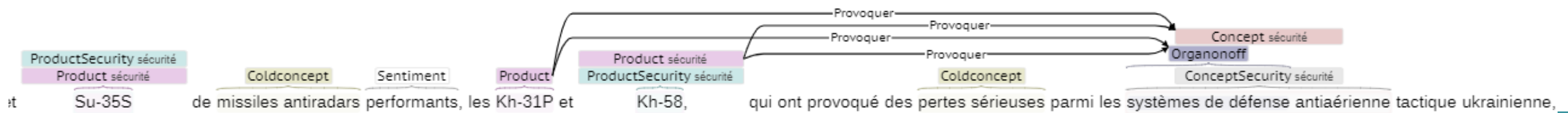
Une croissance d'autant plus intéressante que l'activité est rentable. L'an passé, GeoPost a réalisé 800 millions de profits, deux fois plus qu'en 2019. Mais il aura fort à faire face aux mastodontes du secteur, plus gros et plus connus: Fedex, UPS, DHL. Pour relever le défi, GeoPost va capitaliser sur ses forces: ainsi, en France ou en Pologne, il est le leader de la livraison; en Allemagne et au Royaume-Uni, il est numéro deux; au Brésil et en Afrique du Sud, numéro trois.

«Réservoir de clients»

«Beaucoup de consommateurs qui ont découvert l'e-commerce avec cette pandémie vont continuer à acheter par ce réseau, estime Boris Winkelmann. Et quand on voit que 13 % à 14 % des achats se font sur internet en Allemagne contre 30 % aux États-Unis et en Chine, cela veut dire qu'il y a un vrai réservoir de clients.»

GeoPost va donc croquer le sillon de la livraison aux consommateurs finaux. L'an passé, cette activité a pour la première fois réalisé plus de la moitié du chiffre d'affaires (55 %) du groupe. Cette expansion passera par un développement plus marqué loin de ses bases, avec la volonté de réaliser 20 % de son chiffre d'affaires en dehors de l'Europe en 2025 contre 10 % aujourd'hui.

L'opérateur compte réaliser ce bond en avant avec de la croissance interne mais aussi en rachetant des entreprises. GeoPost en a l'habitude. Ces dernières années, il a pris le contrôle de BRT, le leader de la livraison de colis en Italie, et il est monté à 98 % du capital de Jadlog, son pendant au Brésil.



## EXEMPLES DE CAS D'USAGE



**Automatisation des processus documentaires**, qualification automatique de contenus numériques ou media, classement automatique, cartographies de contenus, ...



**Moteur de recherche sémantique, bases de connaissance, extraction d'indicateurs et de relations** (insight engines)



**Investigations numériques de données textuelles**, forensic, KYC, e-reputation, enquêtes financières, juridiques ou autres, due diligence, analyses de risques, compliance, cybersécurité, ...



**Analyse de mails** entrants, boîtes mail, de messages pour classement, tri, ...



**Analyse des verbatims de tout type** (avis clients, réclamations, expressions spontanées, doléances, livres d'or, etc.) ; **dépouillement d'enquêtes** collaborateurs (QVT) utilisateurs, citoyens, clients, ...



**Web Intelligence** : analyse de l'information web pour extraire des indicateurs et tendances, cartographie d'acteurs, observatoire innovations et start-up ...)

## UNE SOLUTION GÉNÉRALISTE S'ADAPTANT À TOUS CAS D'USAGE DE NOMBREUSES APPLICATIONS VERTICALES

Verticales existantes

- ❖ **Avis clients**
- ❖ **Avis RH**
- ❖ **Moteur de recherche + Tableaux de bord + Sémantique + Relations + Question/anwering**
- ❖ **Knowledge**
- ❖ **Forensic mail**
- ❖ **Analyse de risques business**
- ❖ **Intelligence économique**
- ❖ **Pénalités contractuelles**
  
- ❖ **En phase de POC : renseignement, conformité AFA**

## LES AVANTAGES DE L'IA POUR LE TRAITEMENT DES DONNÉES TEXTUELLES AVEC QWAM TEXT ANALYTICS



### TRAITEMENT EN MASSE ET GESTION DES BIG DATA

Traitement et analyse de grands volumes de documents facilement  
En mode SaaS plateforme hébergée en France sur serveurs QWAM et disponible pour grands volumes  
Installation on-premise suivant besoins clients



### PERFORMANCE ET PRODUCTIVITÉ

Automatisation des tâches de qualification, tri, analyse, organisation, classement ... des documents et contenus  
Amélioration de la productivité et de l'usage des données textuelles



### IA ET ONTOLOGIES

Nombreuses ontologies standards intégrées (Ontologies QWAM)  
Personnalisation des référentiels et vocabulaires en fonction de votre métier (Ontologie Client)  
Combinatoire sémantique/linguistique et machine learning/deep learning/RNN  
Back-office pour pilotage et contrôle humain



### DES USAGES MULTIPLES

API et métadonnées pour applications tierces  
Tableaux de bord d'analyse (pilotage et analyse)  
Moteur recherche ou application d'accès à l'information  
Export de données qualifiées/analysées



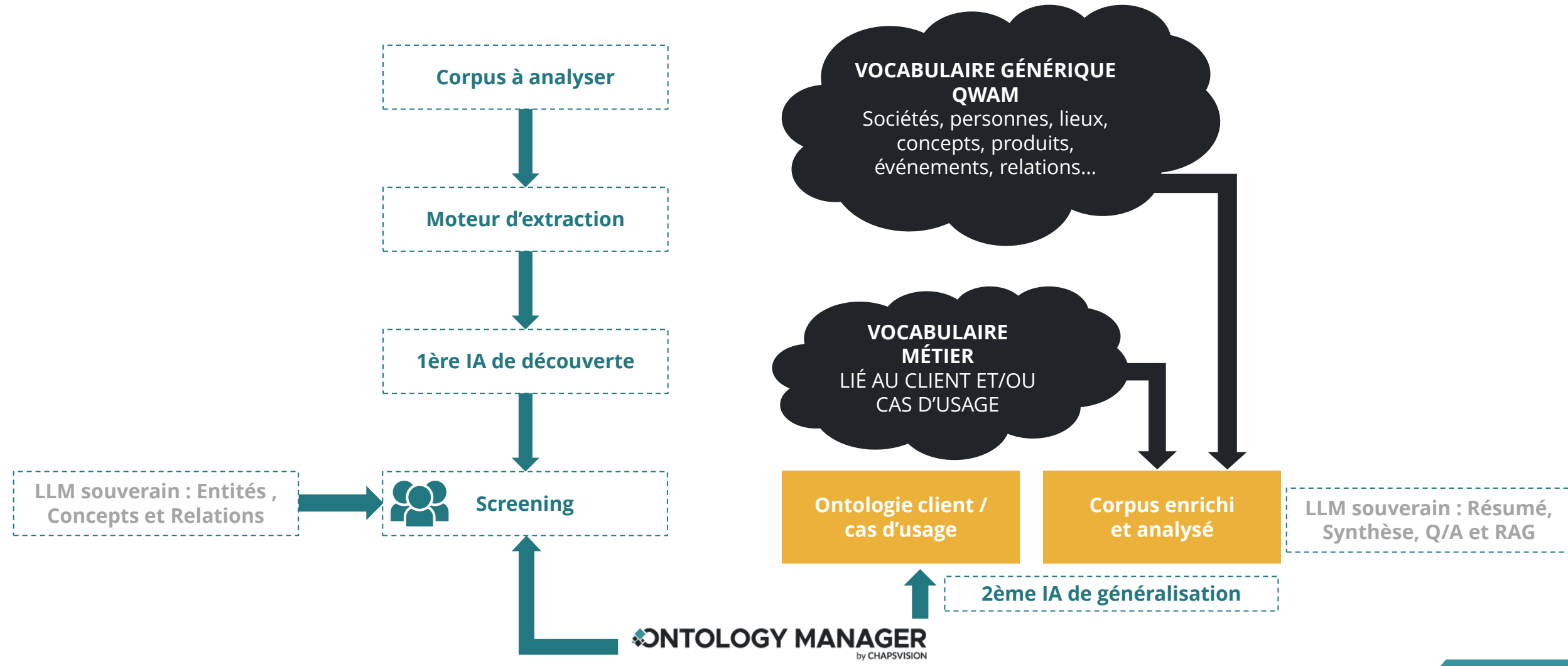
#3

# MÉTHODOLOGIE TEXT ANALYTICS & ADAPTATION AUX CONTEXTES

# QWAM TEXT ANALYTICS

by CHAPSVISION

= Moteur d'extraction + Moteur de découverte + Moteur de suggestions



## PRINCIPE D'UNE CONSTRUCTION INDUSTRIALISÉE D'UNE ONTOLOGIE ADAPTEE A UN CAS D'USAGE

Le module QWAM Ontology Manager permet de construire et piloter les référentiels (Ontologie Client) avec une grande productivité (sémantique + IA).



### **ON NE CONSTRUIT PAS L'ONTOLOGIE/MODELE MOT À MOT**

On réorganise une proposition d'ontologie construite automatiquement  
On rajoute des candidats dans les branches  
→ Processus itératif



### **LES ALGORITHMES À LA RESCousse**

Catégorisation par clustering  
Termes équivalents, contextes équivalents et suggestions par Deep Learning



### **Les LLM ( Large Langage Model) souverains pour augmenter la productivité**

Pré-classification des entités, concepts et relations dans le contexte spécifique



### **ONTOLOGY MANAGER POUR LE PILOTAGE**

Pilotage des référentiels et des regroupements

# UN GESTIONNAIRE D'ONTOLOGIE PERMETTANT DE CONSTRUIRE RAPIDEMENT UNE OU PLUSIEURS ONTOLOGIES METIER

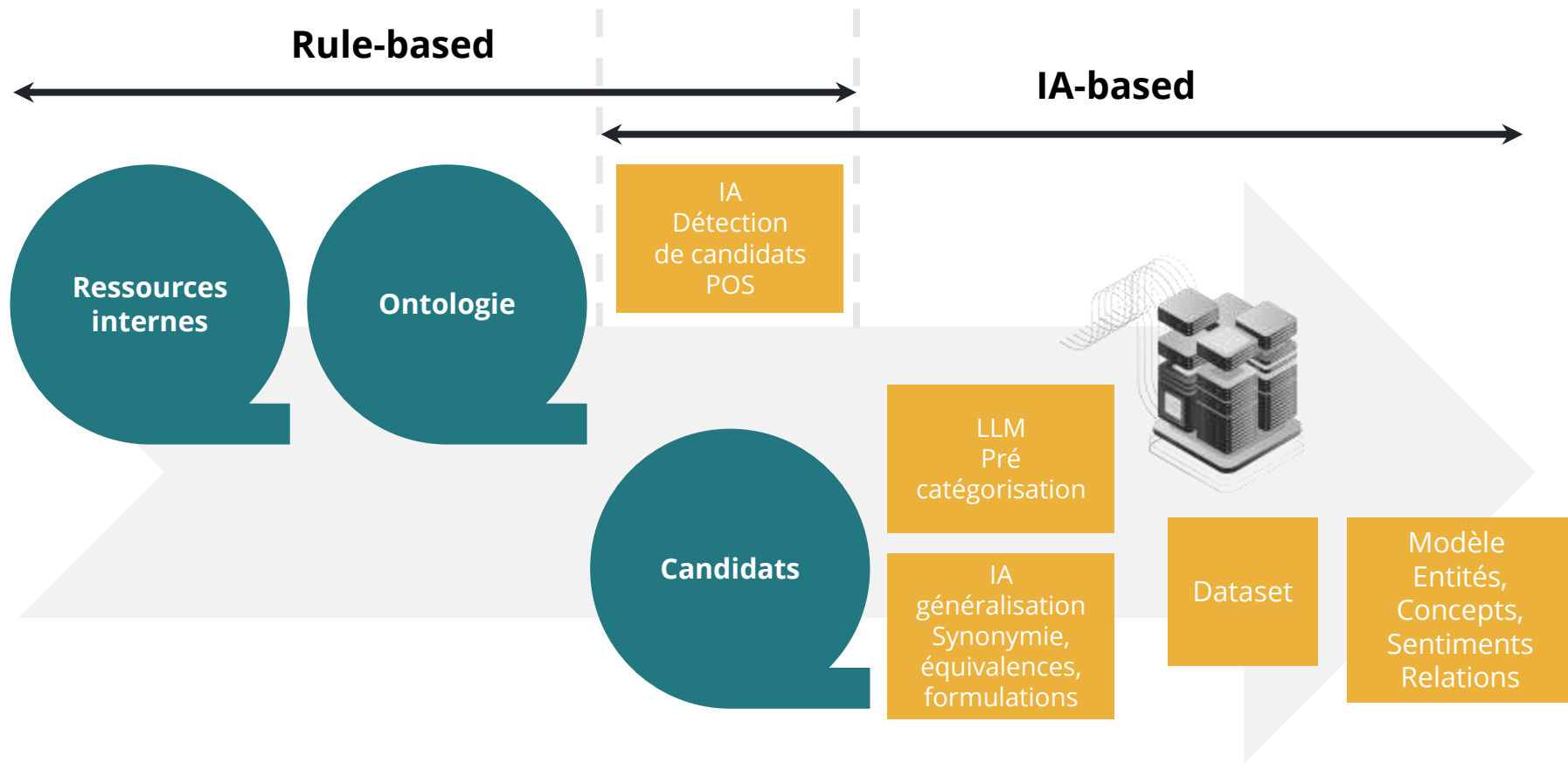
- Construction d'une ontologie assistée par l'IA

The screenshot displays the QWAM Ontology Manager interface. At the top, there is a navigation bar with 'Renseignement', 'ADMINISTRER', and '+ CRÉER' buttons. A search bar contains 'Recherches' and a 'Utilisateurs' icon. Below the navigation bar, there are three main panels:

- Left Panel (Tree View):** A hierarchical tree structure under 'Racine (641)'. It includes categories like 'Corbeille (1)', 'Equipements (82/366)', 'Armes (14)', 'Avions (2/89)', 'Bombardiers (9)', 'Bombes (5)', 'Destroyers (10)', 'Drones (8)', 'Hélicoptères (11)', 'Lance-roquettes multiple (5)', 'Missiles (71)', and 'Navires (13/22)'. Each category has a 'C' icon and a count.
- Middle Panel (List View):** A list of ontology nodes under 'Racine' with a total count of 641. Each node includes a checkbox, a name, and a count. Examples include 'Missile balistique' (119), 'Avion de combat' (102), 'Défense aérienne' (52), 'Système de missiles' (49), 'Équipements militaires' (47), 'Avion de chasse' (40), 'Navire de guerre' (36), 'Système de défense' (33), 'Engin Explosif' (31), 'Véhicules blindés' (31), 'Missile de croisière' (27), 'Système de défense antimissiles' (27), 'Icbm' (25), 'Avion de reconnaissance' (24), 'Matériel militaire' (21), 'Véhicules militaires' (21), and 'Défense anti-missiles' (20).
- Right Panel (List View):** A list of ontology nodes with a total count of 8541. Each node includes a checkbox, a name, and a count. Examples include 'Attentat' (230), 'Exercices militaires' (137), 'Source sécuritaire' (136), 'Régime' (116), 'Sécurité nationale' (85), 'Base militaire' (76), 'Ministre de la défense' (74), 'Prisons' (71), 'Attaque terroriste' (64), 'Source militaire' (63), 'Base aérienne' (62), 'Autorité militaire' (59), 'Frappes aériennes' (57), 'Nouvelle attaque' (57), 'Diplomate' (55), 'Ministère russe de la défense' (53), 'Assassinat' (50), and 'Images satellites' (47).

At the bottom of the interface, there is a toolbar with buttons: '+ NOUVEAU NOEUD | 0', '+ NOUVEAU MOT CLÉ', 'ÉDITER LE GROUPE | 0', 'DÉCATÉGORISER | 0', and 'SUPPRIMER | 0'.

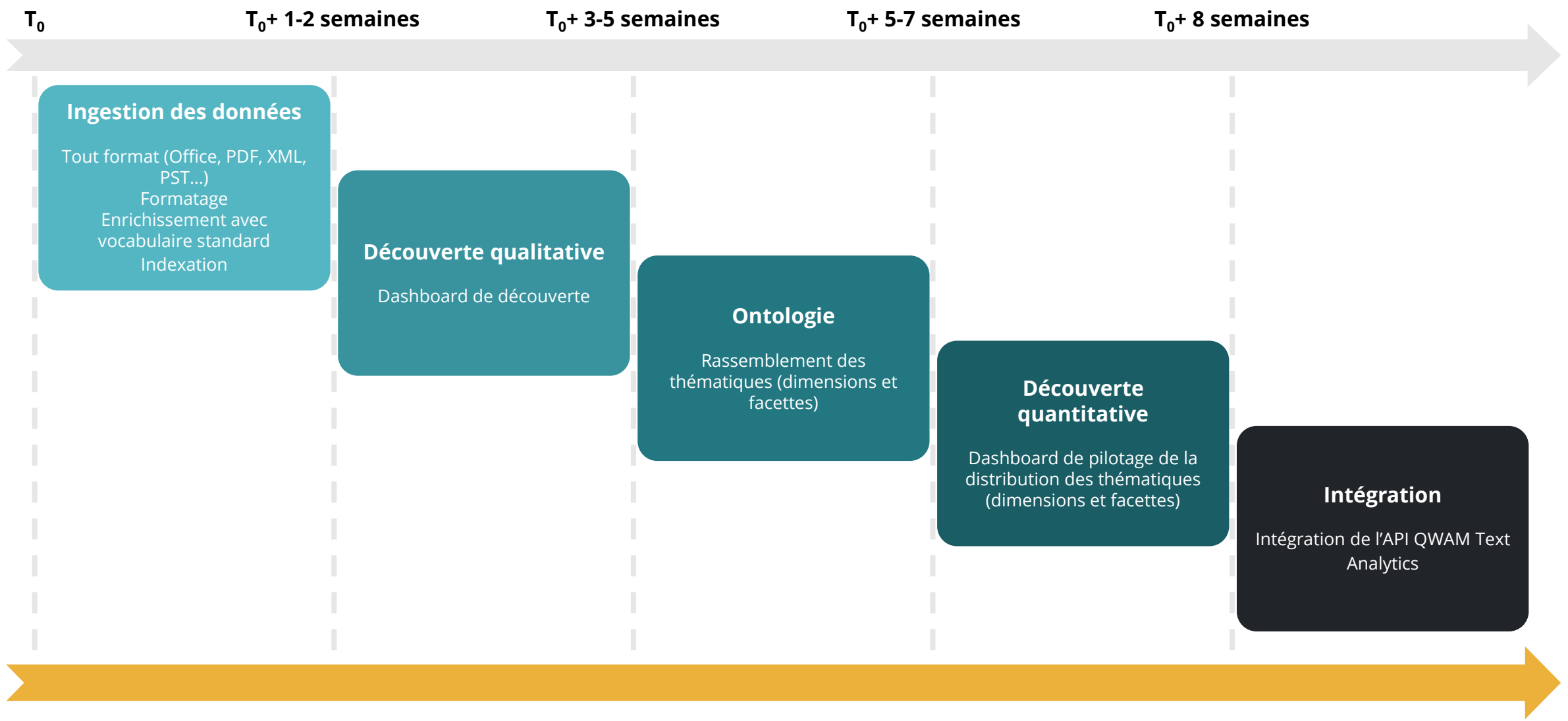
# AVANTAGES D'UNE SOLUTION HYBRIDE : "RULE-BASED" + "IA-BASED" EXPLICABLE // ÉTHIQUE // PEU GOURMANDE EN RESSOURCES



## AVANTAGES

Construire très rapidement une ontologie/modèle sur-mesure  
Éviter l'effet tunnel des projets data

## PHASES D'UN POC/PROJET (EN SAAS)



# QWAM TEXT ANALYTICS

by CHAPSVISION

## Demo API

### Extraits - applications

**Automatisation processus documentaire**  
**Organisation et classement documentaire**

### Analyse de corpus

**Moteur de recherche à facettes  
sémantiques**



#4

## EXEMPLE D'UTILISATION

### Calcul de métadonnées (tagging) de documents pour utilisation dans un système tiers (moteur de recherche, GED, système éditorial, etc.)

- ❖ Back-office de contrôle/pilotage des méta-données calculées automatiquement
- ❖ Utilisation en automatique des méta-données issues de l'enrichissement sémantique/IA
- ❖ Si besoin possibilité de correction/complément manuel

The interface displays a list of tags categorized into four groups:

- Personne:** Sébastien Jean, Basile Lachaud, Laurent Jacobelli, Patrick Artus, Damien Abad, Bruno Le Maire, Gérard Darmanin, Emmanuel Macron, Arnaud Montebourg.
- Personne sujet:** (Empty)
- Société:** Natixis, Renault, Odaxo.
- Société sujet:** (Empty)
- Lieu:** France, Asie, Europe, Chine, Amérique Du Nord, Inde, Seine-Saint-Denis, Bretagne, Poiré, Sochaux, Niort, Charente, Allemagne, Afrique.
- Concepts:** Crise sanitaire, Mondialisation, Concurrence déloyale, Zone euro, Commande publique, Pays émergents, Batteries électriques, Transition écologique, Crise du covid-19, Activité économique, Comptes publics, Produits de santé, Classe politique, Intelligence artificielle, Valeur ajoutée, Innovations de rupture, Politique industrielle, Subventions publiques, Produits Industriels, Domaine sanitaire, Droits de douane, Tarif douanier, Économie de marché, Organisation de la production, Aides publiques.

The screenshot shows a news article titled "Relocalisations : quelles activités faut-il faire revenir en France, et comment ?". A blue overlay on the left side of the article displays the same set of metadata tags as shown in the previous screenshot, demonstrating their application to the document's content.

## EXEMPLE D'UTILISATION

### Les métadonnées calculées sont ensuite utilisées à plusieurs niveaux :

- complément des titres pour aide à la lecture
  - facettes du moteur de recherche
  - classement automatique (dossiers, profils utilisateurs, ...)
  - résumé
- ❖ Back-office de contrôle/pilotage
  - ❖ Utilisation en automatique des méta-données issues de l'enrichissement sémantique/IA
  - ❖ Si besoin possibilité de correction/complément manuel

The screenshot shows a news article snippet from 'Le Figaro Magazine' dated 05/06/2020. The article title is 'Relocalisation, mode d'emploi'. The text discusses the challenges of relocalization in France, mentioning the need for flexibility and the impact of taxes and regulations. Below the text, there are several concept tags categorized by domain: Politics (Préférence nationale, Collectivités locales), Energy (Voiture électrique), Society (Organisation du travail), Business (Prélèvements obligatoires, Made in France, Investissements étrangers), Environment (Voiture électrique), and Transport (Voiture électrique). At the bottom, there are entity tags for names like Xavier Bertrand, André De Chastenay, Renault, and various locations like Allemagne, France, Chine, Bruxelles, Hauts-De-France, and UK.

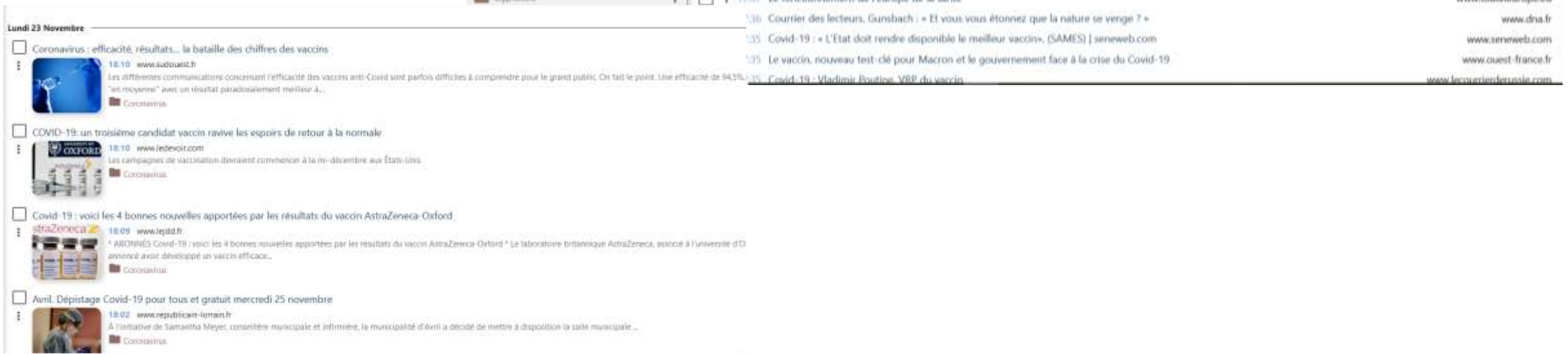
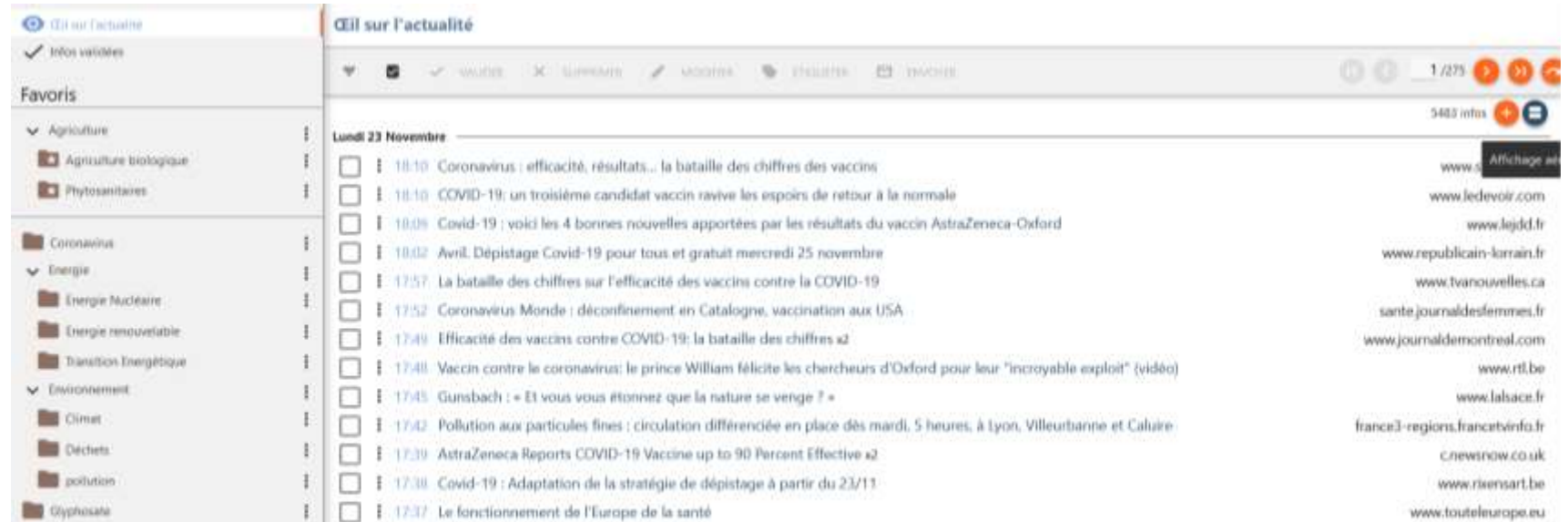
# EXEMPLE D'UTILISATION

## Dualité tableaux de bord/documents

The screenshot displays the QWAM Text Analytics interface, illustrating the duality between dashboards and documents. On the left, a dashboard provides a high-level overview with two donut charts: 'Sources' and 'Personnes'. The 'Sources' chart shows data from various news outlets, with a legend listing: Business Wire, FT Newswire, Home settenews, Radio online, información y noticias de actualidad..., Agencia Europa Press, Seeking Alpha, DW.com, and The Economic Times. The 'Personnes' chart shows data on individuals, with a legend listing: Donald Trump, Emmanuel Macron, Giuseppe Conte, Boris Johnson, Joe Biden, Scott Morrison, and Andrew Young. The main area on the right shows a list of documents, including news articles and market research reports, with a search bar at the top and a 'Tout sélectionner' checkbox.

# CAS DE L'ACCÈS À L'INFORMATION WEB ET DE LA VEILLE WEB

## Automatisation de la sélection d'informations



# CAS DE L'ACCÈS À L'INFORMATION WEB ET DE LA VEILLE WEB

## Multiplication des possibilités d'analyse

Mon cockpit   Mes dossiers   Mes newsletters   Mes sources

### Top des étiquettes

Coronavirus  
Du 16/11 au 23/11

coronavirus vaccin  
vaccin contre le Covid-19  
COVID-19  
Pfizer France AFP  
Astrazeneca essai clinique

### Newsletters

Newsletter - Lundi 28 Septembre 2020

Ask n'Read

Agriculture

**Néonicotinoïdes "tueurs d'abeilles" : histoire d'un échec législatif**  
23/09/2020 - L'Express  
Publié le 23-09-2020 à 08:30 Abonné \* Néonicotinoïdes "tueurs d'abeilles" : Histoire d'un échec législatif \* Faute d'ambition politique et de bonne volonté de la filière sucrière depuis quatre ans, les députés vont assouplir l'interdiction de ces pes

**Vin bio, nature ou HVE... saisissez les différencier**  
23/09/2020 - Le Parisien - Aujourd'hui en France  
Les fabricants de vins bio, nature ou HVE... saisissez les différencier

### Statistiques des infos validées des dossiers

Mes infos validées  
Du 24/09 au 22/11

- Agriculture / Pêche
- Agriculture / Agriculture biologique
- Energie / Énergie nucléaire
- Commerce
- Energie / Énergie renouvelable
- Energie / Transition Énergétique
- Environnement / Climat
- Environnement / Pollution
- Transport et Urbanisme
- Environnement / Océans

### Œil sur l'actualité

Œil sur l'actualité

Lundi 23 Novembre

- 10:48 Déterminés à faire du gaz un accélérateur de la transition énergétique
- 10:43 Réduction des émissions de méthane : Total renforce son engagement avec l'initiative OGMP 2.0
- 10:11 RE2020 : une nouvelle

### Bat et des mots clés

Œil sur l'actualité

Du 17/11 10:08 au 23/11 18:15

- COVID-19
- vaccin contre le Covid-19
- Pfizer
- coronavirus
- Moderna
- France
- AFP
- essai clinique

# TRANSCRIPTION ET ANALYSE D'AUDIO OU VIDÉO POUR INDEXATION AUTOMATIQUE ET RECHERCHE



```

G:\Mon Drive\Marketing...
<Fichier_joint>2501-15029317-536.mp4</Fichier_joint>
<Fichier_joint_size>195520014</Fichier_joint_size>
<IDQES>2501-15029317-536.xml</IDQES>
<Langue>Unknown</Langue>
<Level>2501-15029317-536</Level>
<Level0>2501-15029317-536</Level0>
<Mime>mp4</Mime>
<Plateforme_id>1000</Plateforme_id>
<Titre>Lavf57.56.101</Titre>
<Texte> 0-1.56 Speaker0: bonsoir à toutes et à tous nous attendons vos questions à sa masse internet et réseaux sociaux pour alimenter notre discussion c'est un nouveau coup dur pour les troupes de l'opération barkhane trois soldats sont morts hier tué par un engin explosif au mali dans la zone des trois frontières ce sanctuaire djihadiste où la france avait choisi récemment de concentrer ses efforts effort en huit ans de présence militaire au sahel la france a perdu quarante-sept soldats le président a rappelé ce matin sa détermination à poursuivre la lutte contre le terrorisme sur place ces nouvelles pertes interviennent alors que l'exécutif s'interroge sur les effectifs mobilisés sur place et à la veille d'une visite au tchad du premier ministre la france a-t-elle eu des succès avec l'opération barkhane paraît-il décidé de se désengager sur ce terrain qu'en est-il de la menace terroriste ou plus généralement au sahel terrorisme encore trois soldats tués au sahel c'est le titre de cette émission avec nous pour en parler ce soir vincent je vous êtes journaliste indépendant essayiste vous êtes spécialiste de l'afrrique vous enseigner à sciences po paris je rappelle votre biographie de kadhafi dont la nouvelle édition paraît bientôt en format poche aux éditions Perrin général dominique trinquant vous êtes en chien chef de mission militaire français auprès de l'ONU directeur des relations extérieures de marx et balzac du matériel uniforme de sécu ritter voici menacerait avec nous ce soir vous êtes journaliste à france vingt-quatre spécialiste des mouvements djihadistes enfin euh ni à galets bakayoko vous êtes politologue président de l'african security secteur network vous êtes experte en sécurité en afrique de l'ouest et en afrique centrale bonsoir à tous les quatre merci de participer à ce c'est dans l'air en direct pour revenir juste précisément sur les faits sur ce qui s'est passé voici un astre eux que sait-on des circonstances de ce drame 1.56-113.1 Speaker incertain: on sait que un convoi vercors

```

L'opération barkhane 1400035-1419475 Speaker tah ou'il y en a eu évidemment que les aides et les succès mais après les succès sont de ça reste demeure de succès si vous voulez la tactique parce que si on veut ça de peu en début deux mille vingt d'accord début deux mille vingt il n'y a pas de succès parce que montée en puissance de l'opé islamique en deux mille neuf sur la zone de la zone des trois frontières grosse attaque contre l'armée du sahel grosse attaque contre l'armée malienne grand nombre de victimes son d'ailleurs par exemple qui était un exemple aussi mis en avant comme du retour dans le nord du mali se pas rentrer dans le détail cas attaque donc ça a démontré la montée en puissance de l'opé les techniques pour vendre en novembre deux mille dix-neuf a fait la collision entre deux hélicos à cause de nous passions contre l'état islamique traitement quelques jours avant début novembre un mort et des premières à revendiquer français par l'opé islamique dont quatorze mais en amont donc c'est beaucoup donc il y a eu la réunion de peu ils ont dit on va mettre le paquet sur eux 1419.475-1476.765088 Speaker incertain nombre de trois frontières c'était le cas 1476.765088-1481.517197 Speaker qui parallèlement a fait monter en puissance le rival on va dire de état islamique qui en al qaeda ils ont commencé à se confronter l'été 2014 pour pour résumer ce moment là aussi marquant les cibles prioritaires ce qui on appelle l'état qui des cibles importantes on l'a dit tout à l'heure tous les véhicules tous les anciens les compétences humaines sont du côté al qaeda donc ils n'ont pas renchérit à les taper première exemple l'attaque le plus emblématique troupes djihad donc le patron les djihadistes qui a été tué en juin en plein négociation pour la libération de toureya d'été et qu'on a pénétré en a payé exemple il agit dernier exemple en date avait euh frappes ciblées contre les opérations des forces spéciales au sol porter d'aguer moussa et des lieutenants en qaeda aussi au mali grandes figures au nord mali que de la rébellion touareg deux al qaeda deux succès donc c'est quand il y a eu des personnes comme ça on sort des succès et parallèlement aussi ça en de plus de conflit entre al qaeda vécu avec une idée on comprend bien est-ce que vous nous expliquez avec une idée qui a une surmarché du coup euh entre ces deux organisations est-ce que la question vingt-cinq de =/Texte=

ENCODER LE TEXTE X EFFACER

Afficher plus d'associations Afficher certaines relations

Concepts catégorisés / validés

- Opération barkhane (1000000)
- Présence militaire (1000000)
- Lutte contre le terrorisme (1000000)
- Routes nationales (1000000)
- Épave (1000000)
- Mode opératoire (1000000)
- Attente (1000000)
- Enrôlement (1000000)
- Moyens financiers (1000000)

0-1.56 Speaker0:bonsoir à toutes et à tous nous attendons vos questions à sa masse internet et réseaux sociaux pour alimenter notre discussion c'est un nouveau coup dur pour les troupes de l'opération barkhane trois soldats sont morts hier tué par un engin explosif au mali dans la zone des trois frontières ce sanctuaire djihadiste où la france avait choisi récemment de concentrer ses efforts effort en huit ans de présence militaire au sahel la france a perdu quarante-sept soldats le président a rappelé ce matin sa détermination à poursuivre la lutte contre le terrorisme sur place ces nouvelles pertes interviennent alors que l'exécutif s'interroge sur les effectifs mobilisés sur place et à la veille d'une visite au tchad du premier ministre la france a-t-elle eu des succès avec l'opération barkhane paraît-il décidé de se désengager sur ce terrain qu'en est-il de la menace terroriste ou plus généralement au sahel terrorisme encore trois soldats tués au sahel c'est le titre de cette émission avec nous pour en parler ce soir vincent je vous êtes journaliste indépendant essayiste vous êtes spécialiste de l'afrrique vous enseigner à sciences po paris je rappelle votre biographie de kadhafi dont la nouvelle édition paraît bientôt en format poche aux éditions Perrin général dominique trinquant vous êtes en chien chef de mission militaire français auprès de l'ONU directeur des relations extérieures de marx et balzac du matériel uniforme de sécu ritter voici menacerait avec nous ce soir vous êtes journaliste à france vingt-quatre spécialiste des mouvements djihadistes enfin euh ni à galets bakayoko vous êtes politologue président de l'african security secteur network vous êtes experte en sécurité en afrique de l'ouest et en afrique centrale bonsoir à tous les quatre merci de participer à ce c'est dans l'air en direct pour revenir juste précisément sur les faits sur ce qui s'est passé voici un astre eux que sait-on des circonstances de ce drame 1.56-113.1 Speaker incertain: on sait que un convoi vercors le mur sur la libération nationale le national donc le le titre 113.1-113.04 Speaker0: on sait de vers sur sa

#5

# QWAM TEXT ANALYTICS

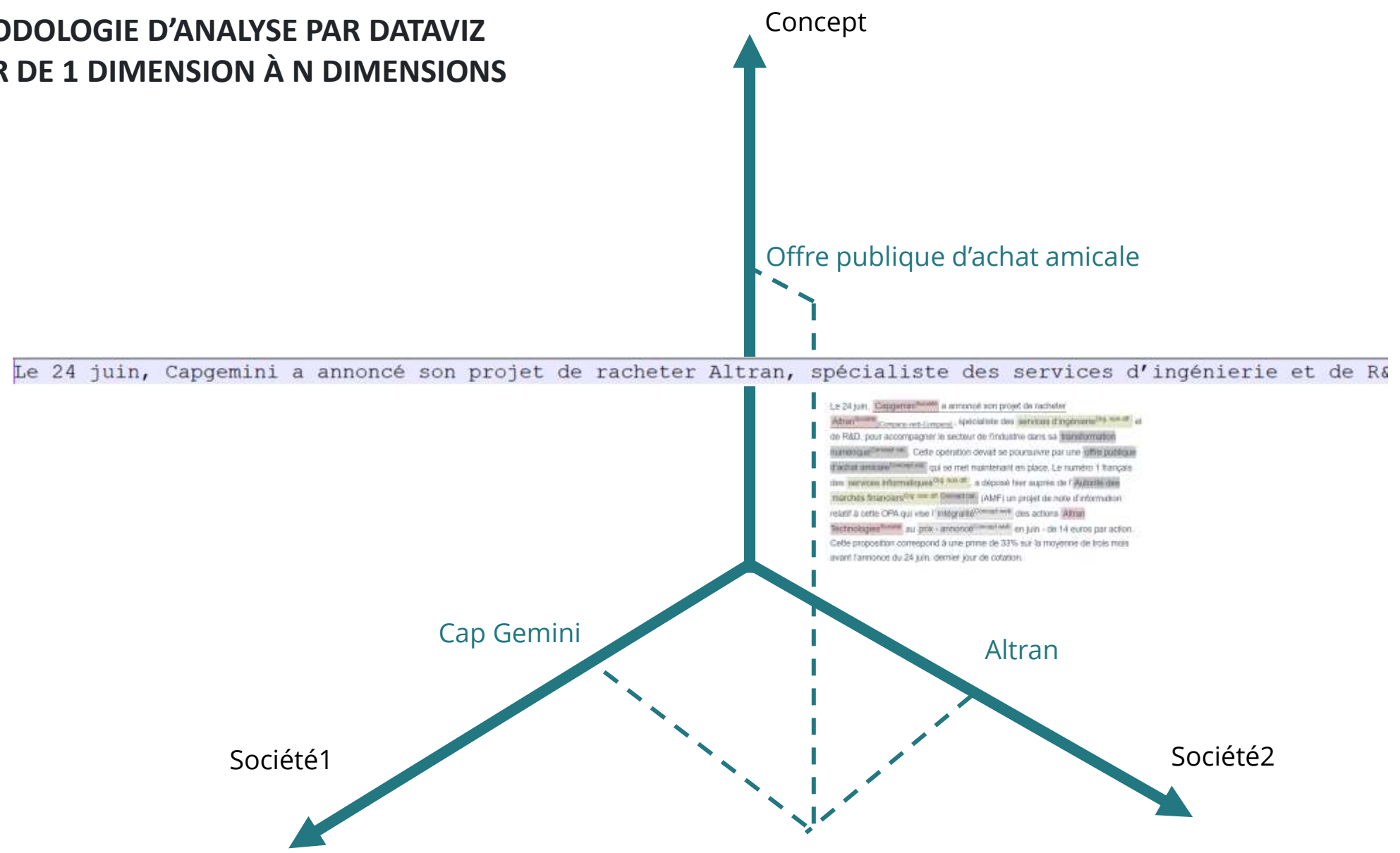
by CHAPSVISION

Analyse de données

Analyse de verbatims

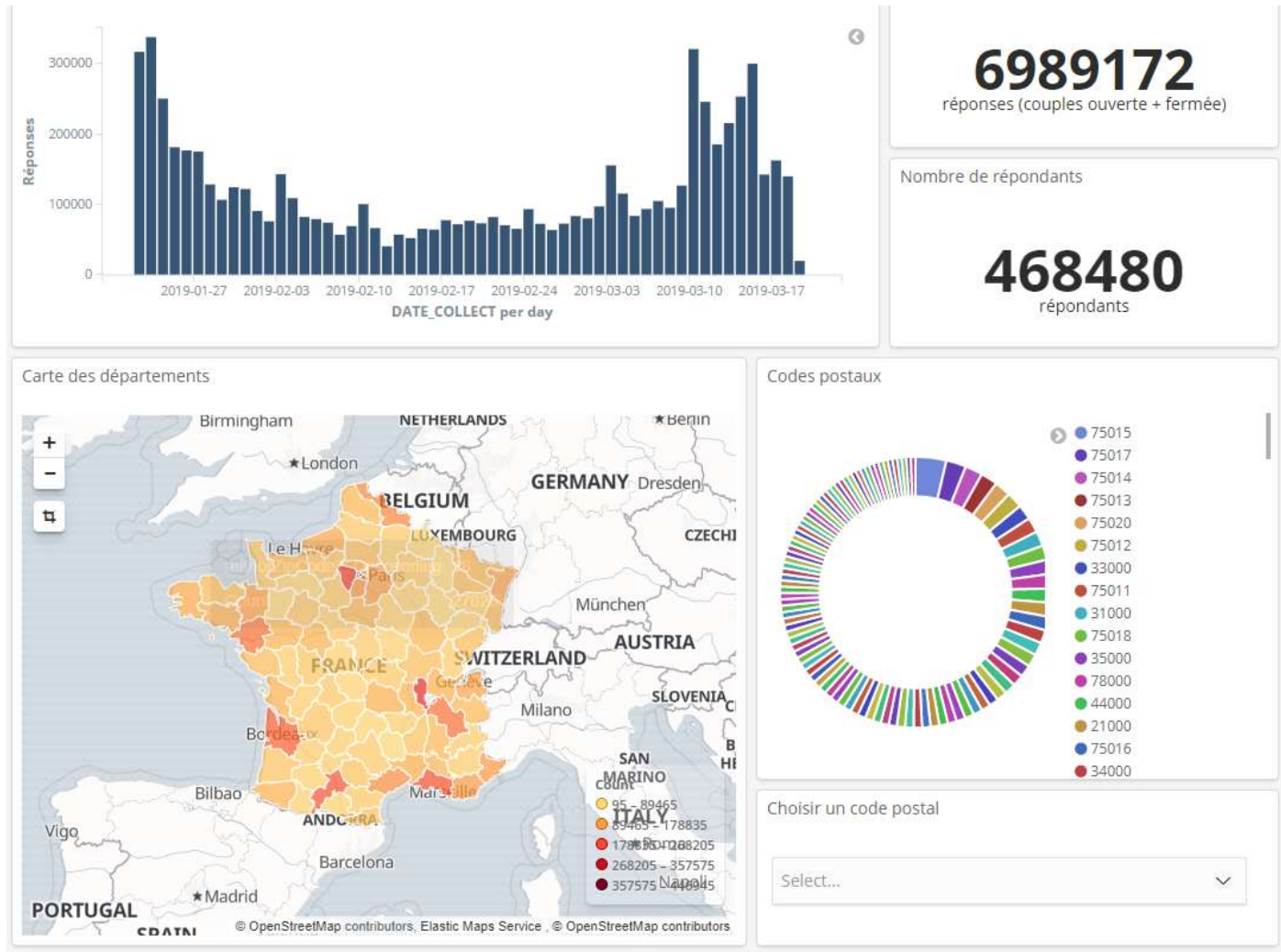
Exploration des data par  
datavisualisation

# MÉTHODOLOGIE D'ANALYSE PAR DATAVIZ PASSER DE 1 DIMENSION À N DIMENSIONS

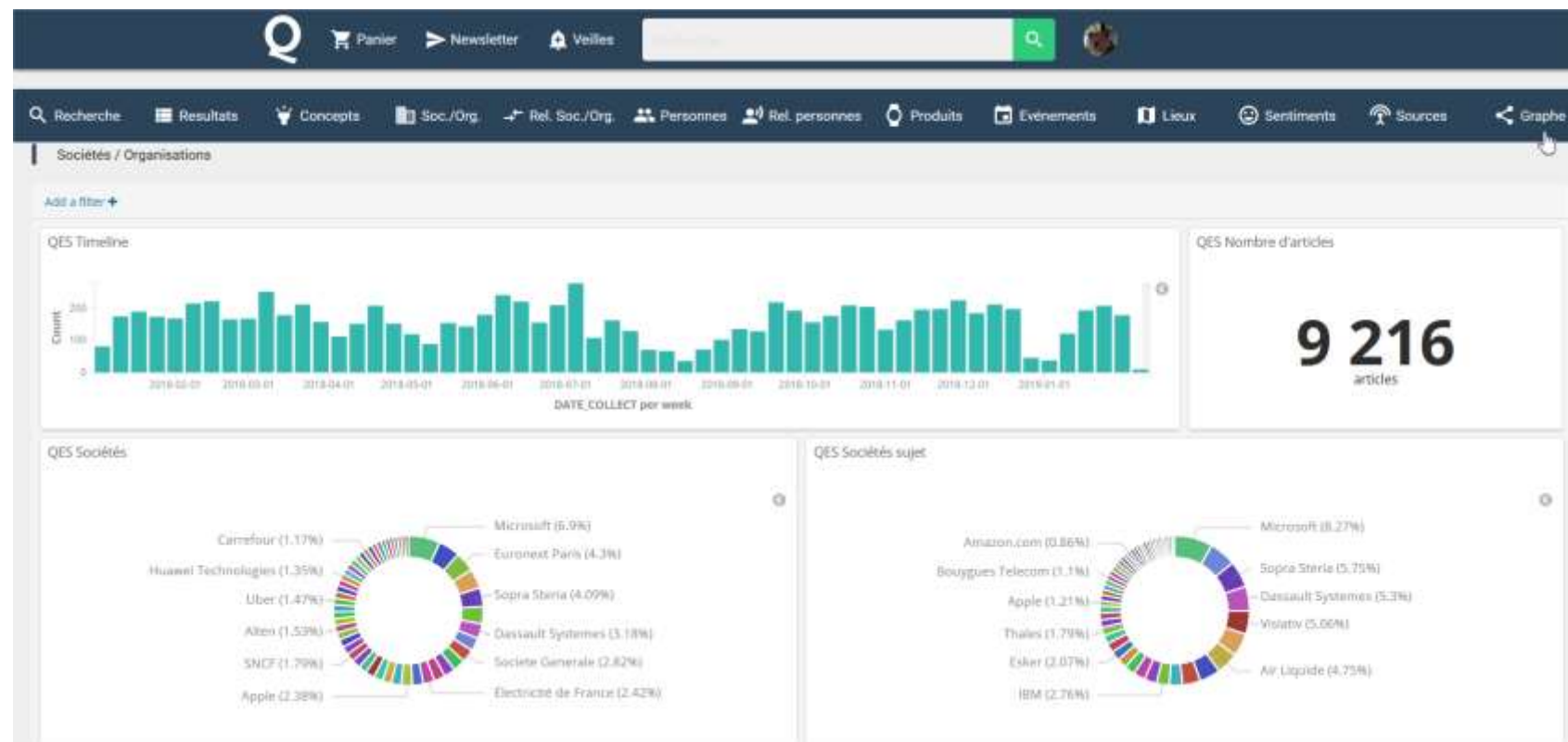




# TOUT REGROUPER DANS UNE DATAVIZ

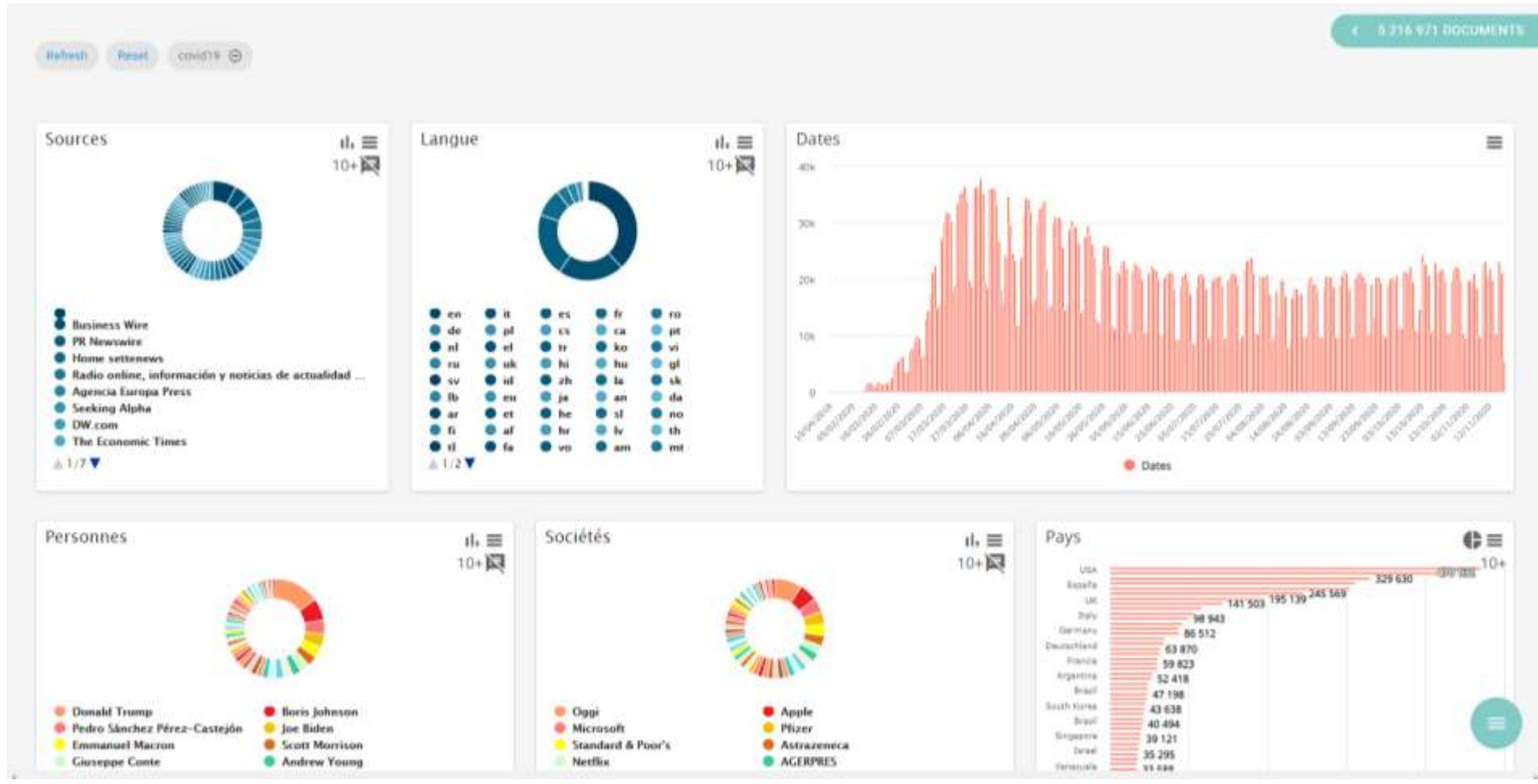


# EXEMPLE D'UTILISATION : TABLEAUX DE BORD INTERACTIFS D'ANALYSE





# EXEMPLE D'UTILISATION : TABLEAUX DE BORD COUPLÉS AVEC RECHERCHE MULTIDIMENSIONNELLE





#6

# LLM / IA GÉNÉRATIVE SOVERAINE

## APPLIQUÉ AU RÉSUMÉ AUTOMATIQUE

## 2 PHASES DE PRODUCTION DU RÉSUMÉ

### Réduire le texte

- **Densité d'entités nommées : Rassembler les phrases les plus denses en entités et concepts**
  - Extractive
  - Basée sur une Ontologie
- **Clustering d'idées forces : Rassembler les phrases de sens approchant**
  - Extractive
  - Clustering des phrases similaires BERT

### Générer un résumé

- IA Générative, par tranches
- LLM : ChatGPT (pas souverain) , ou LLama2, Mistral, OpenChat, Gemini

# RÉDUCTION PAR DENSITÉ D'ENTITÉS NOMMÉES

## Avantages

- Spectre d'analyse fin contrôlant le résumé
- Adapté aux contenus **explicites**
- Bon contrôle de la taille
- Résumé « **de qui et de quoi parle-t-on** »

## Inconvénients

- Nécessite une ontologie
- Résumé un peu « haché »
- Ratio de compression limité si pas de spectre



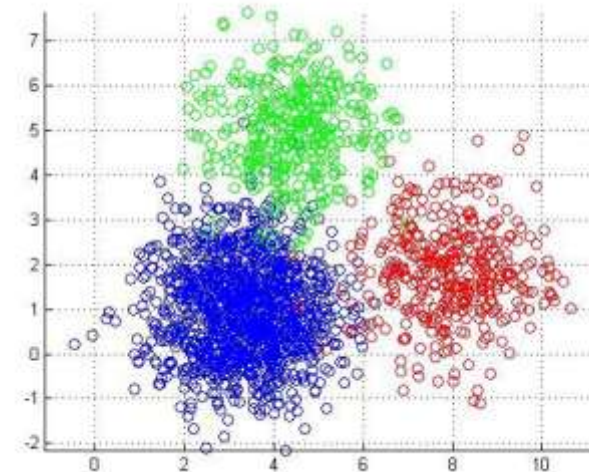
# RÉDUCTION PAR CLUSTERING D'IDÉES FORCES

## Avantages

- Adapté aux **contenus longs**
- Bon contrôle de la taille
- Résumé des **idées forces du texte**

## Inconvénients

- Résumé pas toujours fluide
- Pas toujours tous les noms et entités



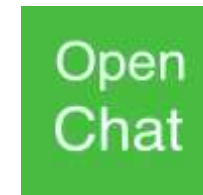
## RÉSUMÉ PAR IA GÉNÉRATIVE

### Avantages

- Adapté aux **contenus argumentaires ou complexes**
- Résumé **fluide**

### Inconvénients

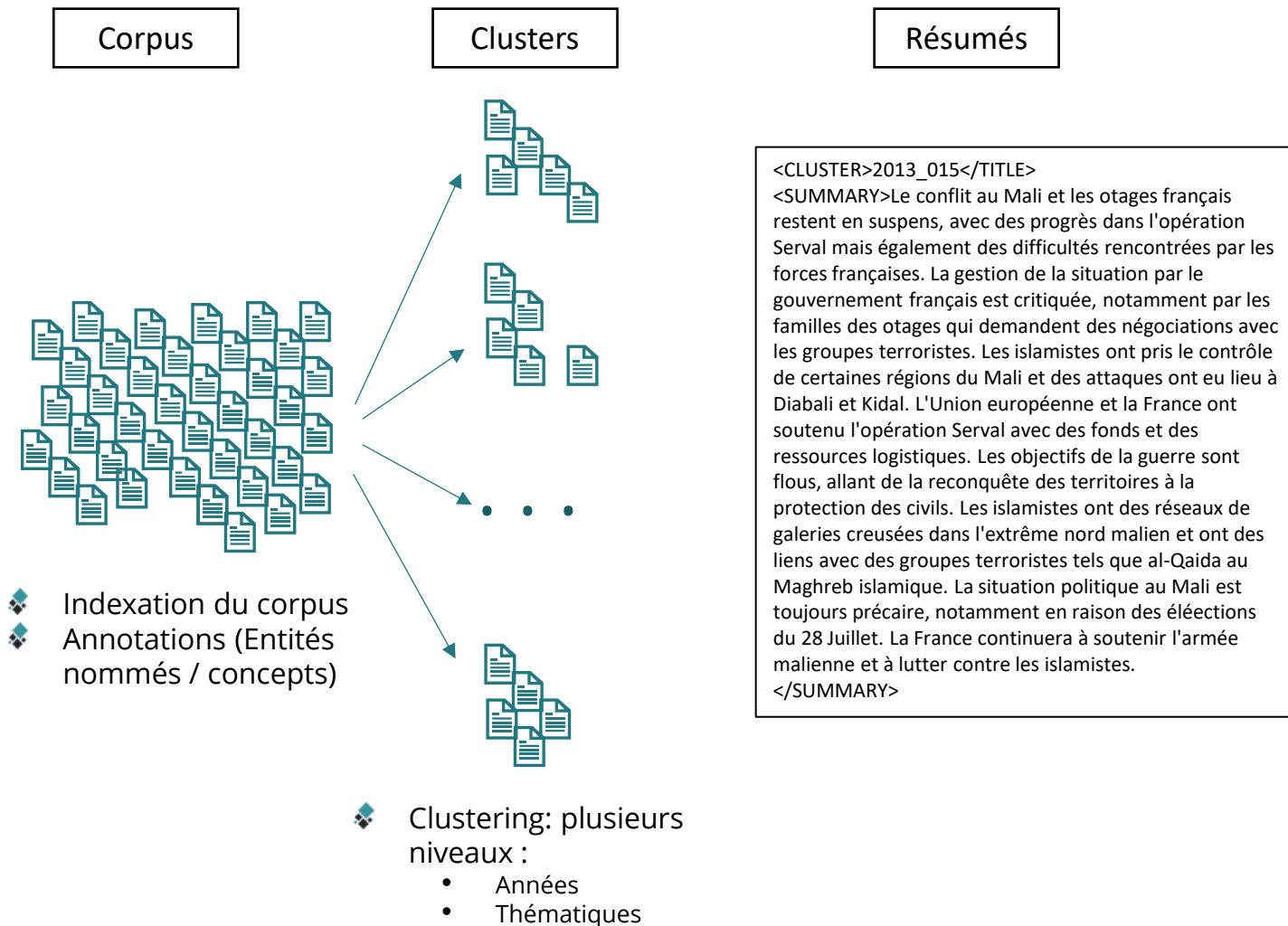
- Taille difficile à contrôler,
- Ratio de compression limité
- Hallucinations possibles



## RÉSUMÉ PAR SPECTRE D'ANALYSE

- Contenu explicite
- Le résumé doit être guidé
- Le contenu des sections est contrôlé
- Plusieurs axes de résumé (Ex: **rapports d'expertise de sinistres**)
  - Les intervenants
  - Les dégâts
  - Le sinistre (conséquences)
  - Les responsabilités
  
- Taille normalisée
- 30 pages réduites à 30 lignes

## GÉNÉRATION DE RÉSUMÉS / SYNTHÈSES DE PLUSIEURS TEXTES



- ❖ Tri et hiérarchisation des documents :
  - ❖ Par année
  - ❖ Par thématiques
  - ❖ 30 clusters / année
  - ❖ 1 à 10 sous-cluster par cluster
- ❖ Réduction du contexte aux éléments les plus informatifs :
  - ❖ Entités Nommées (personnes, organisation, lieux, dates, etc..)
  - ❖ Idées-forces et concept
- ❖ Génération de résumés grâce au LLM
  - ❖ Llama 2 et Mistral
  - ❖ Définition d'une taille de sortie (nombre de mots)
  - ❖ Choix de la langue

#7

# FONCTIONNEMENT DU RAG

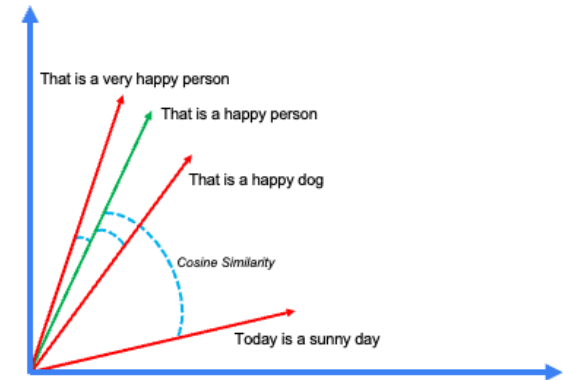
# Embeddings

This is a very happy person

0.235	0.585	0.789	0.123	0.005	0.057	0.887	0.895	0.694	0.889
-------	-------	-------	-------	-------	-------	-------	-------	-------	-------

He is a happy guy

0.235	0.585	0.124	0.123	0.666	0.057	0.887	0.789	0.694	0.889
-------	-------	-------	-------	-------	-------	-------	-------	-------	-------





Chunks

Embedding


Espace latent

# Chunking

 per header + per table with headers

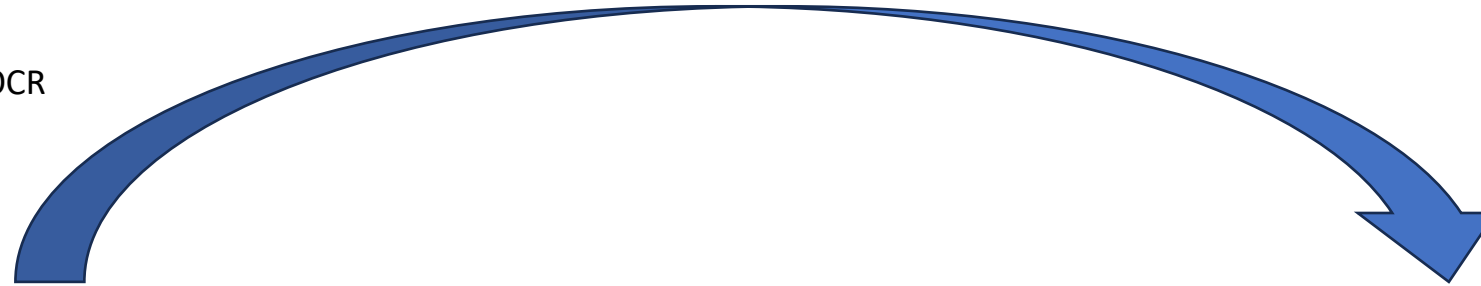
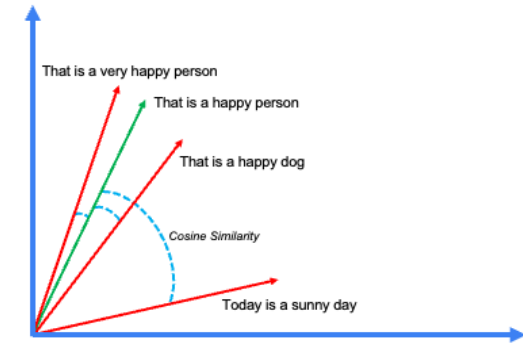
 per page + OCR if image

 per slide

 per sheet + per table with headers

 OCR  


- A la phrase
- Récursif
- Structuré



  
elasticsearch



Documents

Chunking

Strategy

Embeddings

Indexing



# Le process du RAG



Quels sont les rapports entre le **président de la République** et les membres de la **junte militaire au Mali** ?

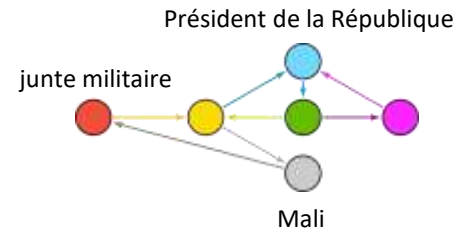
Routeur de stratégie

Question answering  
Step back, reformulation, Parent document  
TOT, COT, GOT

Expansion de requêtes

Qui est le président de la République ?  
Qui sont les membres de la junte militaire Malienne ?  
Quelles sont les relations entre la France et le Mali ?  
Quels sont les conflits entre les gouvernements de France et du Mali ?

Traversée de graphes



Contextes



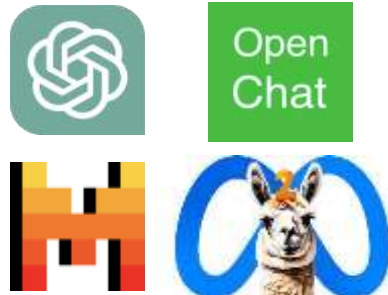
Reranker



Réponse

Les rapports entre le gouvernement français et la junte militaire au Mali se sont tendus pour plusieurs raisons :  
1. Le président français a déclaré ....

LLM

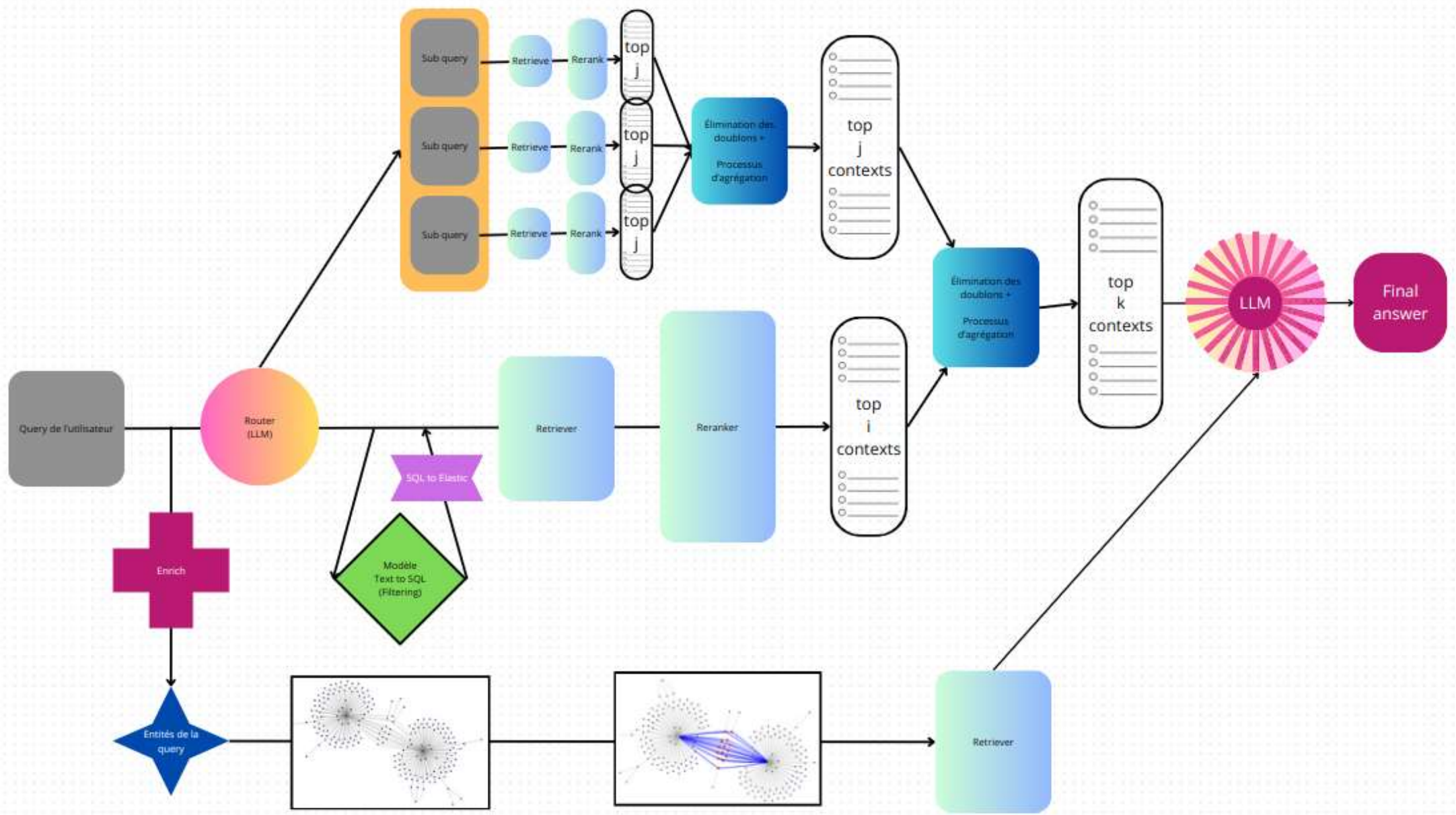


Prompt

You must answer the question...



# ARCHITECTURE DU RAG



# AGRÉGATION ET ANALYSE DE DONNÉES TEXTUELLES ET MEDIA

## MOTEUR DE RECHERCHE ET D'EXPLORATION

## FRONT-OFFICE (APPLICATION) QWAM INFOMEDIA

## QWAM NLP AVEC SPEECH-TO-TEXT ET TRADUCTION INTÉGRÉE



## LES TYPOLOGIES DE DONNÉES TEXTUELLES DANS LES ORGANISATIONS



Fichiers et documents  
textes, contrats, rapports,...



Applications métiers et  
documentaires



Informations web &  
Réseaux sociaux



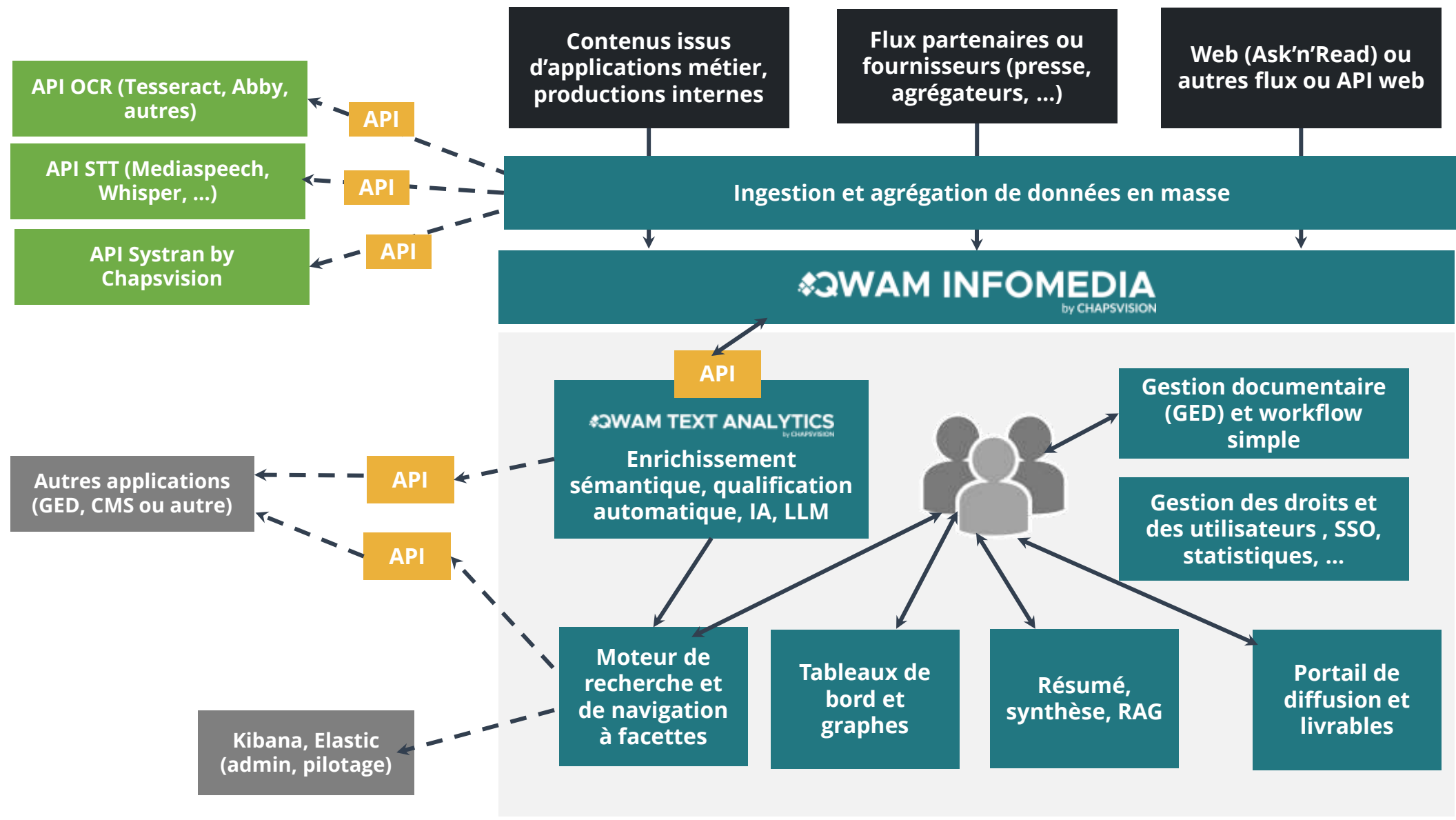
Emails



E-commerce, plateformes  
clients



Multimédia (transcriptions  
vidéo/audio, images)



## EXEMPLES DE CAS D'USAGE



**Automatisation des processus documentaires**, qualification automatique de contenus numériques ou media, classement automatique, cartographies de contenus, ...



**Moteur de recherche sémantique, bases de connaissance, extraction d'indicateurs et de relations** (insight engines)



**Investigations numériques de données textuelles**, forensic, KYC, e-reputation, enquêtes financières, juridiques ou autres, due diligence, analyses de risques, compliance, cybersécurité, ...



**Analyse de mails** entrants, boîtes mail, de messages pour classement, tri, ...



**Analyse des verbatims de tout type** (avis clients, réclamations, expressions spontanées, doléances, livres d'or, etc.) ; **dépouillement d'enquêtes** collaborateurs (QVT) utilisateurs, citoyens, clients, ...



**Web Intelligence** : analyse de l'information web pour extraire des indicateurs et tendances, cartographie d'acteurs, observatoire innovations et start-up ...)

## LES FAMILLES D'APPLICATIONS



### API QWAM TEXT ANALYTICS

- API pour l'extraction de méta-données de documents issus ou gérés dans d'autres applications



### API QWAM TEXT ANALYTICS + STACK ELK

- API pour l'extraction de méta-données de documents issus ou gérés dans d'autres applications
- Tableaux de bord Kibana pour pilotage, consultation et analyse



### QWAM TEXT ANALYTICS (extraction + résumé + RAG) + QWAM INFOMEDIA

- ❖ Applications fonctionnant avec l'ontologie standard :
  - ❖ Presse, média et contenus généralistes
  - ❖ Veille et IES générale
  - ❖ Exploration/découverte de contenus (dump téléphone, disques durs, etc.)
- ❖ Applications fonctionnant avec une ontologie verticale faite par QWA% by Chapsvision
  - ❖ CTI
  - ❖ Leaks
  - ❖ Extraction d'indicateurs d'entreprises (business insights)
  - ❖ Autres en cours
- ❖ Applications nécessitant une ontologie spécialisée
  - ❖ Projets avec ontologie client







# DUALITE TABLEAUX DE BORD ET MOTEUR DE RECHERCHE (NAVIGATION A FACETTES, RECHERCHE MULTI ET MONO CHAMP) – FONCTIONS COMPLEMENTAIRES (SELECTION, WORKFLOW, ALERTES, PORTAIL, ...) – RESUMES – RAG

The screenshot displays the QWAM by CHAPSVISION interface, which is a dual-view dashboard and search engine. The top navigation bar includes the logo, menu items like 'Découverte', 'Relations', 'Résumés', and 'Veille web', and a search bar with the placeholder 'Rechercher dans le full text'. Below the navigation, the interface is split into two main sections.

**Left Panel (Dashboard):**

- Dates:** A line chart showing data trends over time from 2011 to 2020.
- Lieux:** A horizontal bar chart showing the number of documents for various locations. The data is as follows:
 

Lieu	Nombre de documents
Mali	10+
France	39 365
Sahel	17 719
Pers	16 619
Niger	14 351
Bamako	10 611
Afrique	8 707
Burkina-Faso	8 707
Afghanistan	8 616
Libye	8 616
Tchad	8 616
Europe	8 616
USA	7 727
Algérie	7 727

**Right Panel (Search Results):**

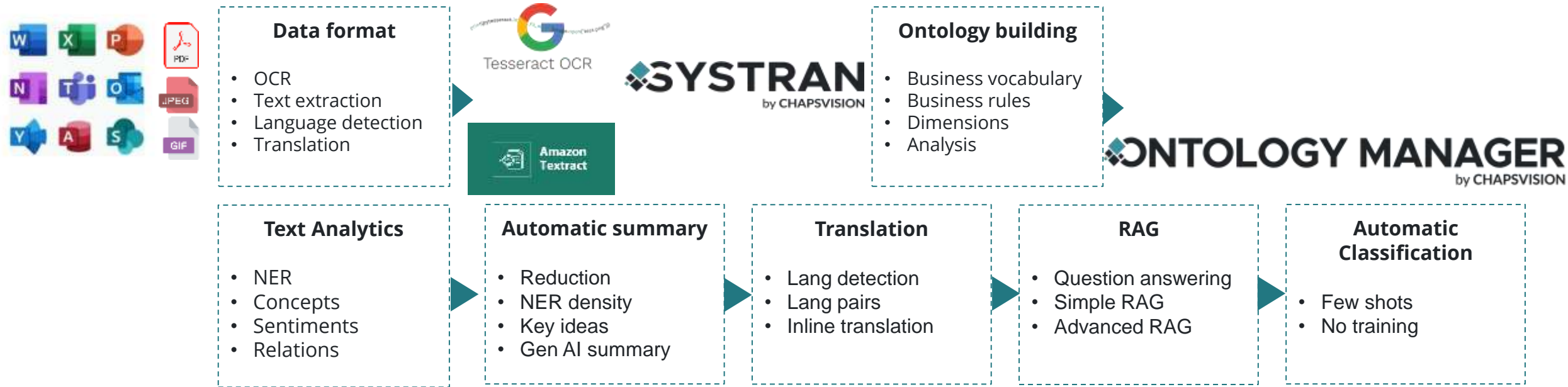
- Search Query:** "Qui est le chef historique de Boko Haram ?" with 20 results.
- Result 1:** "Le chef historique de Boko Haram est Aboubakar Shekau." with a checkbox for selection and sorting options.
- Result 2:** "Le rapt des lycéens déstabilise le Nigeria". The snippet reads: "Six ans après l'enlèvement de 276 lycéens, vendredi, dans la ville de Kankara, d'une majorité de ses combattants...". The date is 17/12/2020 - Ouest France Rennes Nord. The snippet continues: "Patrick ANGEVIN. Au moins 330 au cœur de celui du président Buhari. leader historique du groupe Boko Haram lycéens, vendredi, dans la ville de Kankara majorité de ses combattants en 2016. C de son bastion, montre qu'il n'en est rier d'éradiquer Boko Haram. Le message po".
- Faceted Search:** A horizontal bar at the bottom of the right panel shows various filters like 'USA', 'ALGERIE', 'MALI', 'SOUTIEN LOGISTIQUE', etc., with a total count of 286.
- Result 3:** "2013 | France combat terrorisme au Mali et au Nigeria : 7 otages français libérés, mais menaces continues". The snippet reads: "nigeria, boko haram, francais, cameroun, otages, famille, sept, france, islamiste, nous, ansaru, collomp, lenlevement, deux, dune, enlevés". It includes a 'Level 1 | 369 documents | 0,37%' summary and a 'Resume\_2013/Resume\_2013\_001.xml' link.
- Summary Chart:** A donut chart showing the distribution of documents across different categories.



#10

# STACK TECHNIQUE

# STACK TECHNIQUE



## QWAM TEXT ANALYTICS by CHAPSVISION

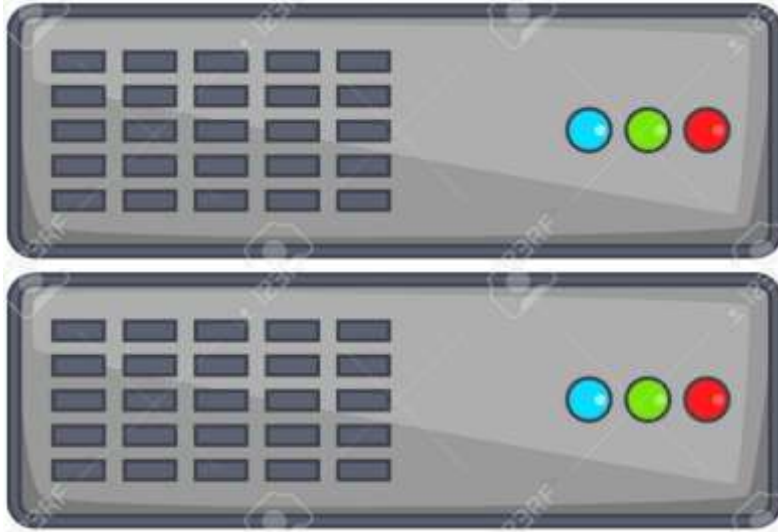


QWAM Models

QWAM Vocabularies

# SERVEURS

**Process** Debian  
48 Threads, 128GB RAM, 4x4TB **SSD**



**App + Kibana** Debian  
32 Threads, 64GB RAM, 4x4TB **HDD**



**Translate** RHEL/Rocky  
32 Threads, 64GB RAM, 4x2TB **HDD**



**GPU**  
48 GB

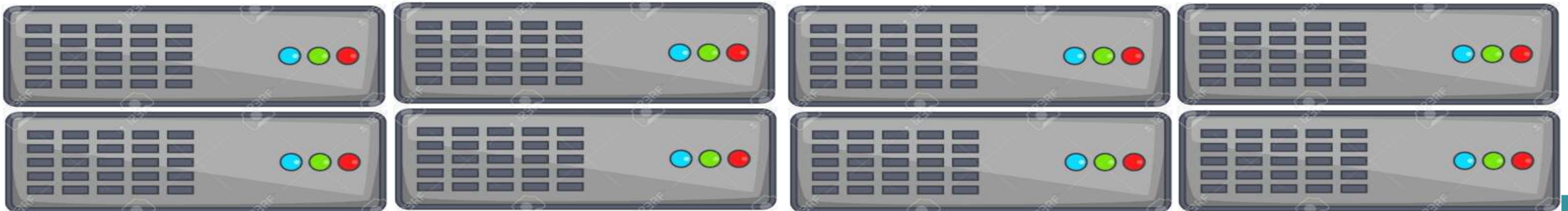
Modèles IA annexes



Modèle LLM



**Elastic** Debian  
32 Threads, 64GB RAM, 4x4TB **SSD**



# CONFIGURATION MINIMALE POUR 1 ENVIRONNEMENT

**Process**  
48 Threads, 128GB RAM, 4x4TB **SSD**



Debian

**App + Kibana**  
32 Threads, 64GB RAM, 4x4TB **HDD**

Debian

**Translate**

32 Threads, 64GB RAM, 4x2TB **HDD**

RHEL/Rocky



**GPU**  
48 GB

Modèles IA annexes

Modèle LLM



**Elastic**  
32 Threads, 64GB RAM, 4x4TB **SSD**

Debian



A man with a mustache is looking intently at a computer screen. The screen displays various data visualizations, including line graphs, bar charts, and a grid of code. The overall aesthetic is dark and technical, with a blue and black color palette. The text 'CHAPSVISION DATA MAKE SENSE' is overlaid on the screen in white. Below the main text, the phrase 'MERCI DE VOTRE ATTENTION' is written in a large, bold, white font. There are also some decorative elements like circles and lines on the screen.

**CHAPSVISION**  
DATA MAKE SENSE

**MERCI DE VOTRE ATTENTION**