

Enrichir la connaissance des utilisateurs des données du site insee.fr

Le recours à l'analyse textuelle



- 1 . - Contexte et objectifs**
- 2 . - Analyse textuelle : principaux enseignements**
- 3 . - Résultats – exemple de proto-persona**

01 CONTEXTE & OBJECTIFS

- Contexte : mise en place de nouveaux services à destination des utilisateurs avancés des données
- But : Compléter la recherche utilisateur
 - Focus sur les besoins en termes d'accès à la donnée
 - Les attentes des utilisateurs



Fournir des outils synthétiques permettant de dresser le profil des utilisateurs cibles et leurs attentes



- Méthode initiée par Alan Cooper (1999) – phase fondamentale d'une démarche de conception centrée utilisateur
- Le persona est une fiche présentant un archétype d'utilisateurs-type

Deux phases de construction :

- « recherche primaire »
- « recherche secondaire »

Full Name Kense

Goals

- The goals this user hopes to achieve.
- A task that needs to be completed.
- A life goal to be reached.
- Or an experience to be felt.

Frustrations

- The frustrations this user would like to avoid.
- The obstacle that prevents this user from achieving their goals.
- The problems with the solutions already available.
- The product or service that currently does not exist.

Bio

The bio should be a short paragraph to describe the user journey. It should include some of their history leading up to a current use case. It may be helpful to incorporate information listed across the template and add pertinent details that may have been left out. Highlight factors of the user's personal and/or professional life that make this user an ideal customer or your product.

Personality

Introvert ☐ Extrovert ☐

Analytical ☐ Creative ☐

Conservative ☐ Liberal ☐

Passive ☐ Active ☐

Motivations

Incentive ☐

Fear ☐

Achievement ☐

Growth ☐

Power ☐

Social ☐

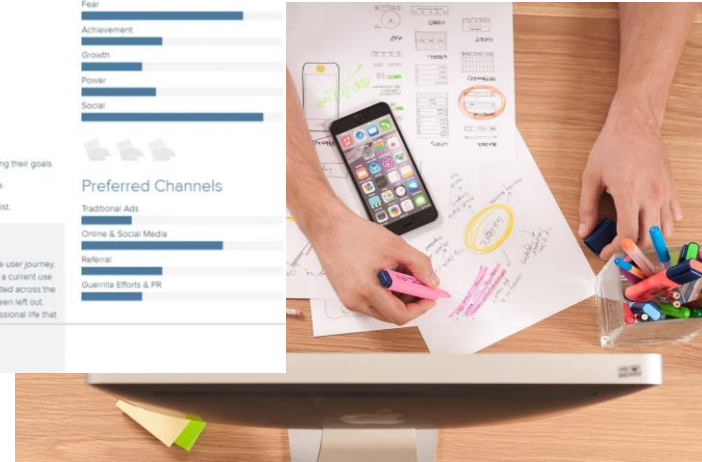
Preferred Channels

Traditional Ads ☐

Online & Social Media ☐

Referral ☐

Quantita Efforts & PR ☐





Utile à toute l'équipe de
conception

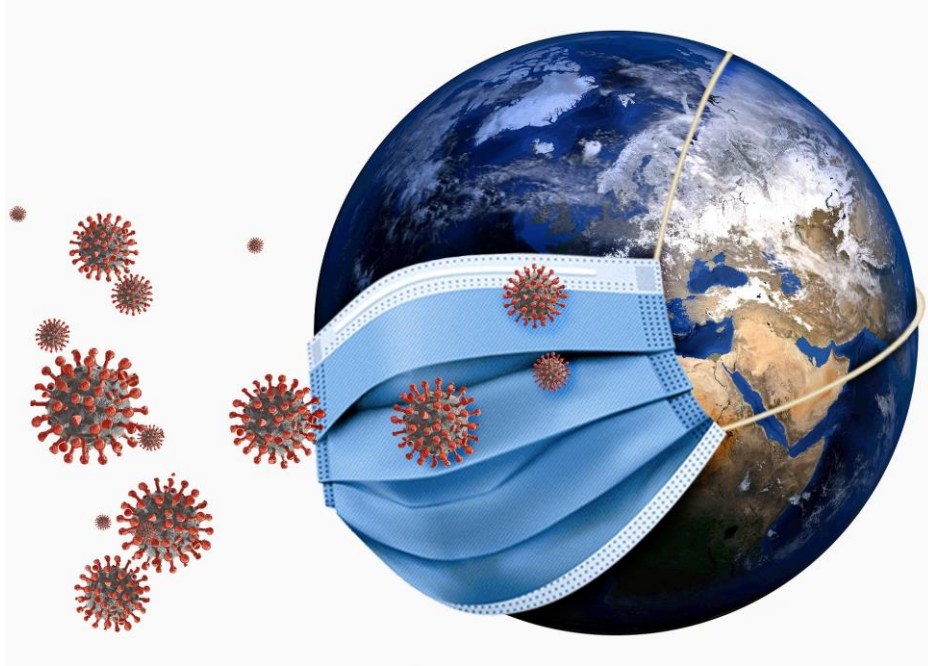
Mobilisable pour des projets
connexes

Matériau pour des brainstorming

Facilite la communication

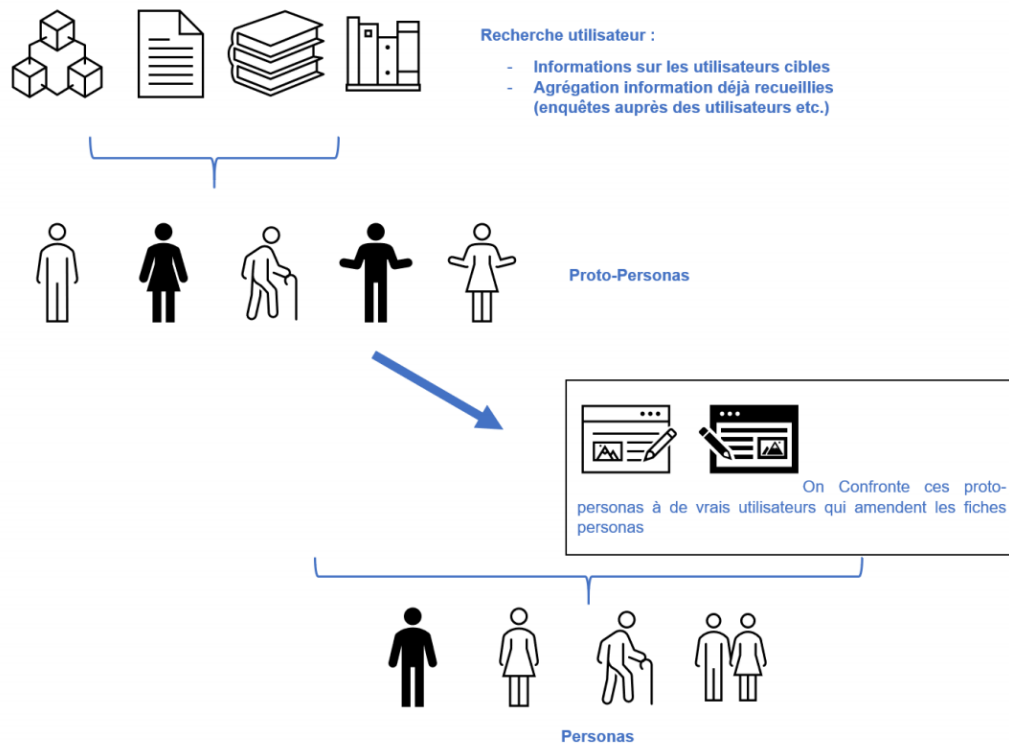
Utilisable avec d'autres outils

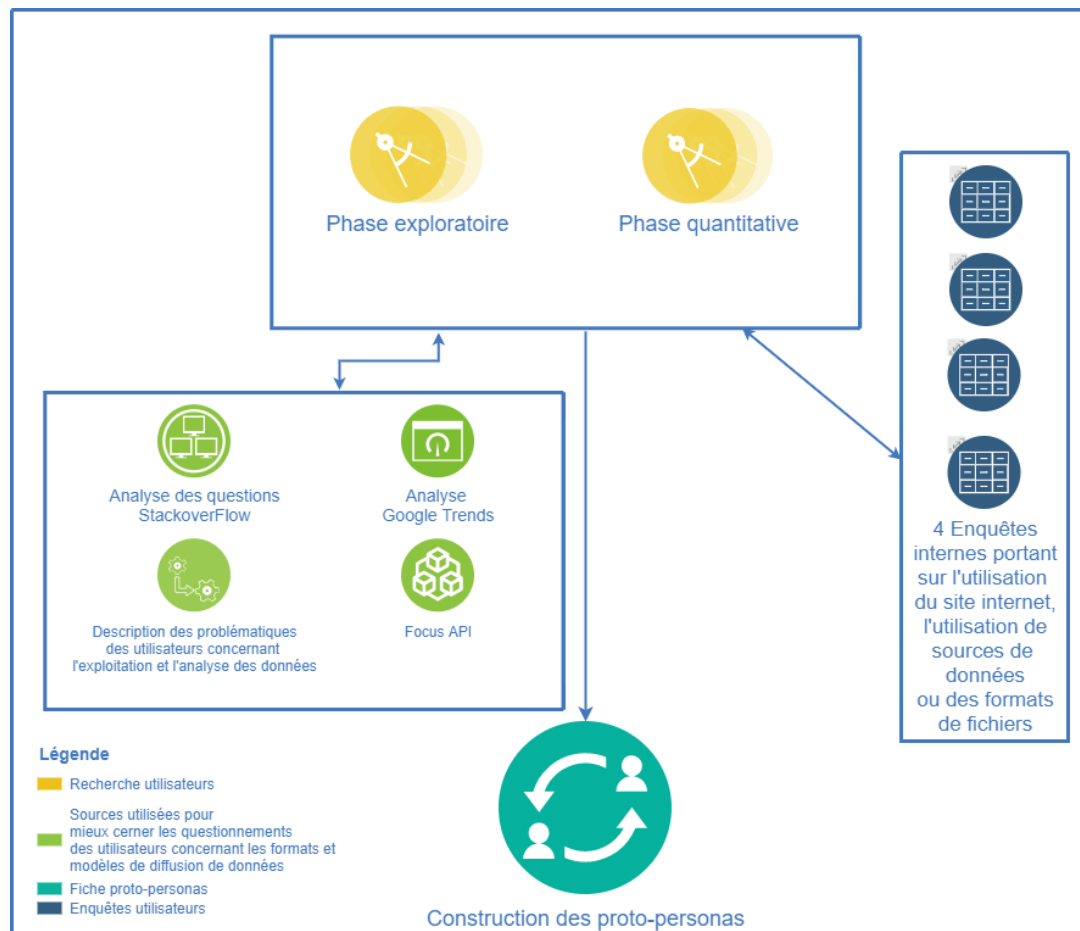
....



- Recours à la méthode des proto-personas
 - Méthode alternative en « mode guérilla »
 - Moins consommatrice de ressources

CONSTRUIRE DES PROTO-PERSONAS



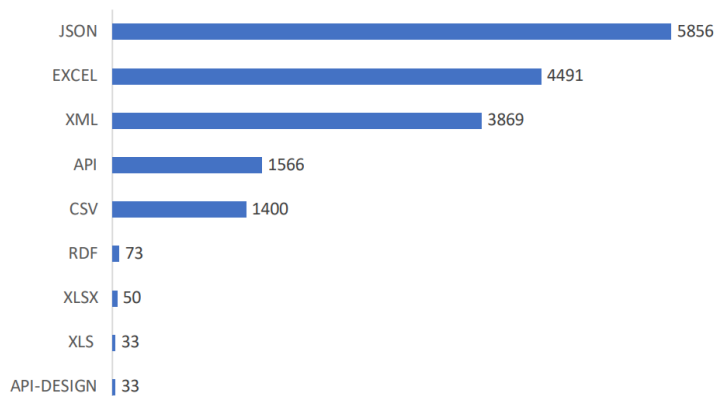


02 ANALYSE TEXTUELLE

- Focus sur deux exploitations particulières
 - Analyse questions Stackoverflow
 - Analyse enquête auprès des utilisateurs du site insee.fr

- Stackoverflow est une communauté ouverte d'entraide entre codeurs
- Questions portant sur les mots clés :
 csv – api – api-design
 json – rdf – xlsx
 xls – xml – excel

Nombre de pages sur le site stackoverflow.com (50 questions par page)

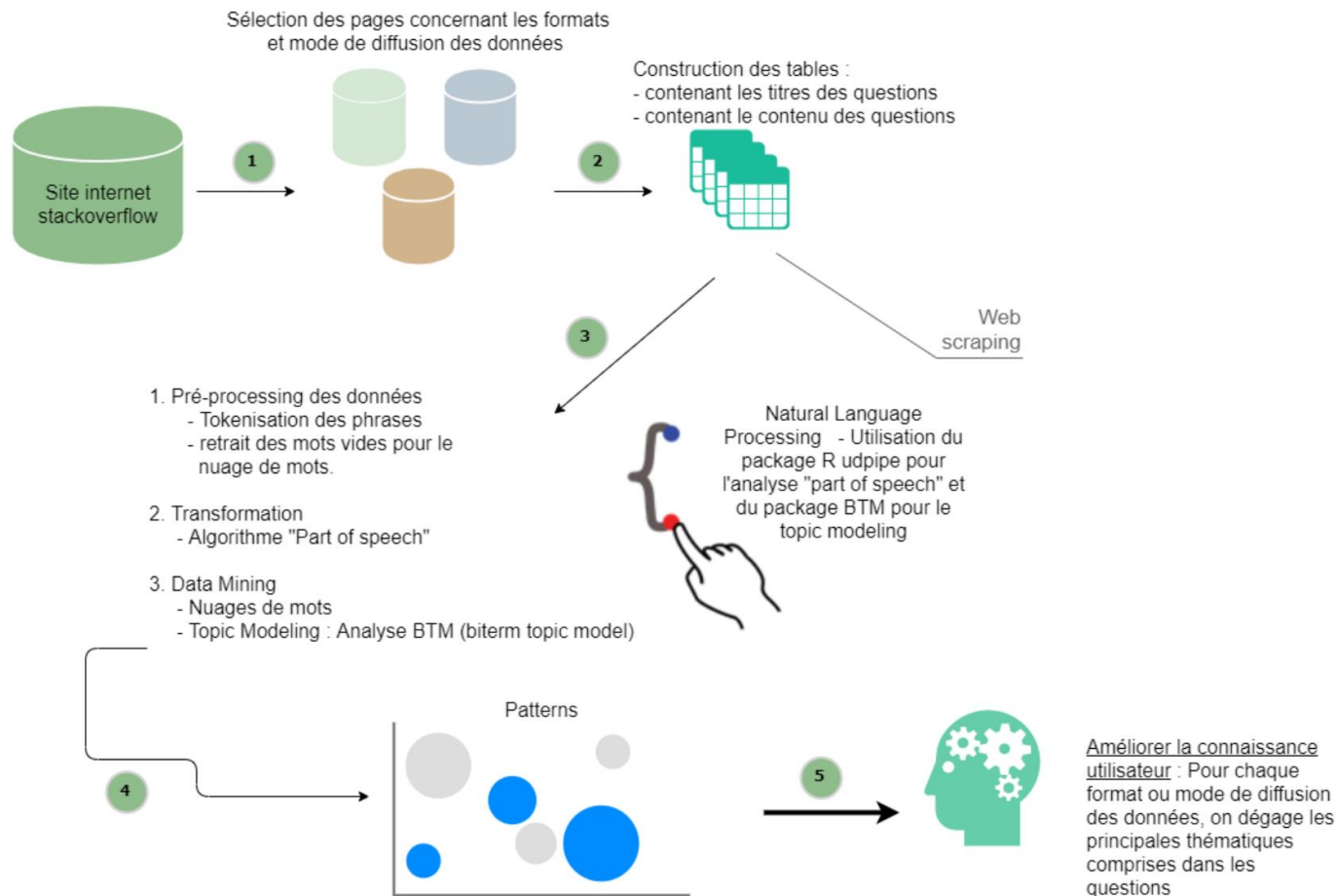


Source : stackoverflow.com (fin mai 2020)

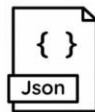
Mot-clé	Nombre de pages existantes sur le site Stackoverflow	Nombre de pages scrapées	Nombre de verbatims présents sur le site	Nombre de verbatims récupérés
RDF	73	72	3 650	3 600
XLSX	50	50	2 500	2 500
XLS	33	33	1 650	1 650
EXCEL	4 491	500	224 550	25 000
API	1 566	500	78 300	25 000
API-DESIGN	33	33	1 650	1 650
JSON	5 856	500	292 800	25 000
XML	3 869	500	193 450	25 000
CSV	1 400	500	70 000	24 978
Total	17 371	2 688	868 550	134 378

Source : stackoverflow.com (fin mai 2020)

ANALYSE DES QUESTIONS STACKOVERFLOW



JSON



Plusieurs clusters de thématiques :

- Des questions portant sur la lecture ou l'importation de fichiers JSON dans d'autres outils (tels que R par exemple)
- Des questions relatives aux API (utilisation API Google, Facebook...)
- Des questions sur la structuration JSON(key/value)
- Des questions sur des codes erreurs retournés par l'utilisation d'API

XML



Plusieurs clusters de thématiques :

- Des questions portant sur l'utilisation de web services
- Des questions portant sur le scraping ou le parsing de fichier XML pour les utiliser via des packages R ou Python
- Des questions portant sur le protocole SOAP (des demandes de précision sur les requêtes)
- Des questions relatives à la structure d'un document XML (XPath, node etc.)

RDF



Plusieurs clusters de thématiques :

- Des questions relatives à la conversion de fichiers RDF/XML au format JSON-LD
- Des questions portant sur l'utilisation de Jena (API Java pour RDF)
- Des questions sur l'utilisation du langage de requête SPARQL
- Des questions sur la syntaxe RDF et sur des explications sur le concept de web semantic



Informations à conserver pour la création des proto-personas



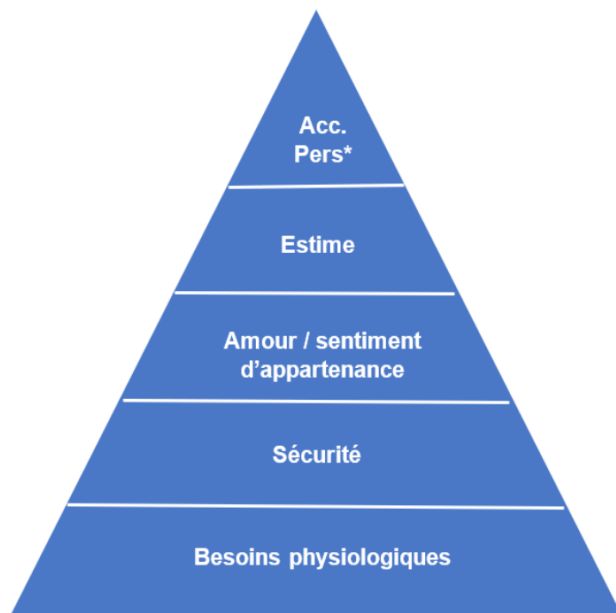
- Pour les Expérience Maps, les informations concernant l'utilisation des APIs ne seront pas simplement utiles aux profils développeurs mais également pour les profils datascientist.
- L'importance du groupe de questions portant sur les APIs des géants du numériques tels que Google, Facebook, Amazon, Twitter... conduit à regarder la structuration de ces APIs ainsi que leurs usages par les développeurs ou datascientists.
- CSV et Excel répondent aux besoins de plusieurs personas → voir si ces formats couvrent l'ensemble de leurs besoin

- Enquête auprès des internautes ayant visité le site et répondu à l'enquête
- Note de satisfaction globale puis opinion sur les points forts et faibles du site

Analyse sur deux questions

- « D'APRÈS VOUS, QUEL EST LE POINT FORT DU SITE ? »
- « D'APRÈS VOUS QUEL EST LE POINT FAIBLE DU SITE ? »

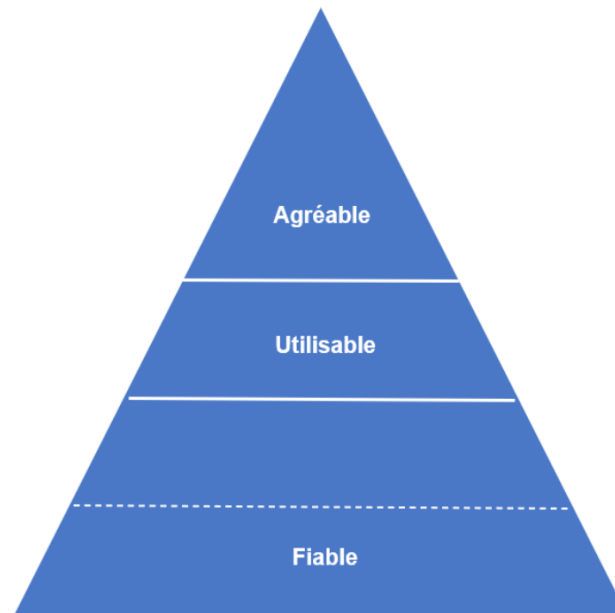
Pyramide des besoins – Maslow



* Accomplissement personnel

Source : Walter A, (2017, 5ème ed. 2017). Design Emotionnel

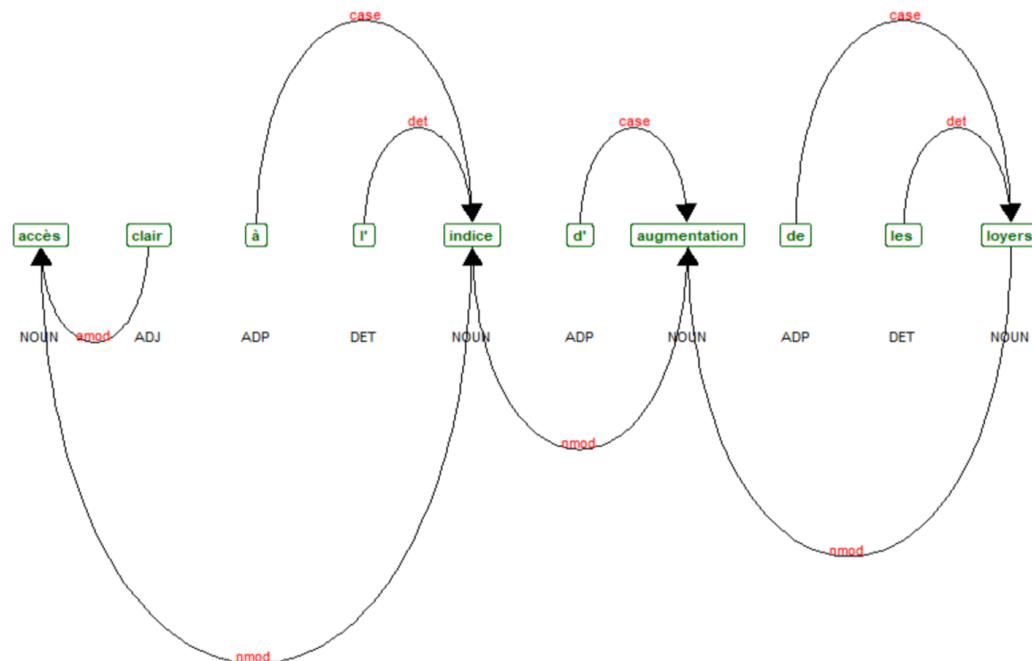
Pyramide des besoins d'un utilisateur d'un service numérique – Walter



Exemple : tokenisation + analyse
part-of-speech + analyse des
relations sur une réponse d'un
inseenaute à la question sur les
points forts du site

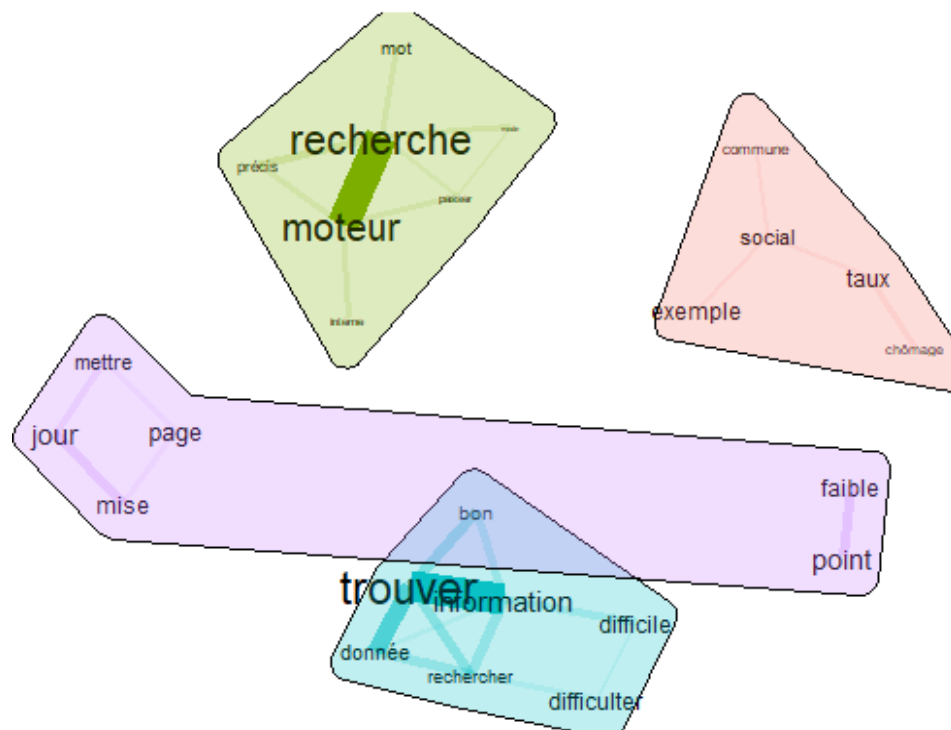
Dependency Parser

tokenisation, parts of speech tagging & dependency relations



Topics des points faibles du site (1890 verbatims)

Biterm topic model



03 RÉSULTATS

UN EXEMPLE DE FICHE PROTO-PERSONA



Kévin

25 ans, Data analyst

Secteur d'activité : privé

Ses outils :



Mode de diffusion des données

API ●●○○○

Tableur ●●●●●

CSV / TXT ●●●●●

RDF ○○○○○



Biographie –

Après avoir obtenu son diplôme d'ingénieur, Kevin a débuté sa carrière en tant que Data analyst. C'est son premier emploi et il le dit lui-même « c'est le meilleur métier du monde ». A la fois technique, statistique et consultant, Kevin se considère comme un évangéliste de la donnée. Sa mission consiste à répondre à une problématique et trouver les techniques permettant aux métiers de s'emparer de ses analyses. Il s'adapte sans cesse à tout type de public, néophytes ou plus aguerris. Il met à disposition ses analyses via des tableaux de bord ou des rapports.

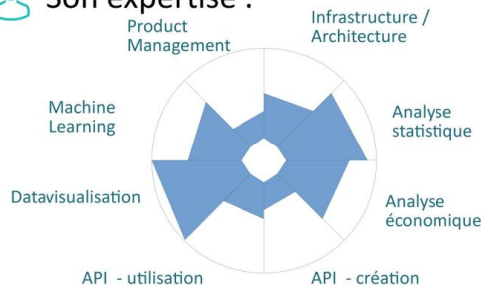


Ses objectifs :

- Récupérer des données permettant de répondre à la problématique métier
- Formater et nettoyer les données (internes et externes) pour qu'elles soient exploitables
- Mettre en forme les résultats de ses analyses sous forme de datavisualisations interactives
- Mettre à disposition des métiers les résultats des analyses



Son expertise :



Ma recherche de données est rythmée par les demandes métiers et les analyses à faire pour y répondre

Types d'usage données Insee

Découverte

Consultation

Exploration

Analyse

Partage information

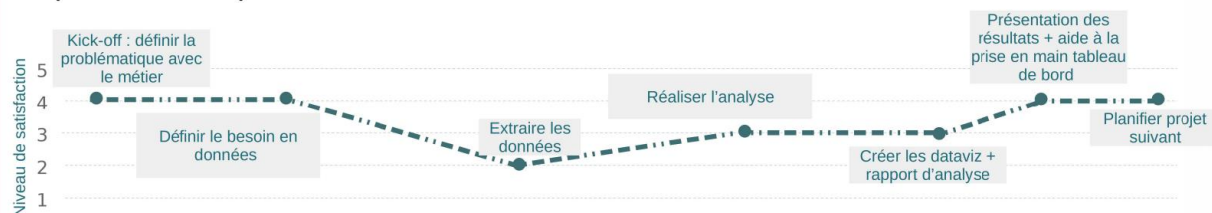
Intégration SI interne

Rediffusion

Ses critères de jugement

- Avoir un panorama des sources de données pouvant répondre à sa problématique
- Aide pour formater les données (dictionnaire des codes etc.)
- Inspiration pour restituer et vulgariser l'information

Experience Map de Kévin



ÉTAPE	1. Échanger avec la direction métier afin de définir la problématique data	1. Définir les étapes du workflow technique	1. Récupérer les données	1. Simplifier la prise en main des résultats	1. Poursuivre les efforts dans la stratégie data-driven de l'entreprise
Objectifs	2. Traduction du besoin – se faire une idée de la faisabilité		2. intégrer les données externes à l'analyse	2. Permettre la mise à jour des résultats des analyses	
Opportunités	<ul style="list-style-type: none"> Occasion d'échanger sur les données internes avec les métiers Explicitier la demande et proposer des améliorations 		<ul style="list-style-type: none"> Découvrir les données : en tirer des enseignements pour compléter l'analyse Créer de la documentation Récupérer des dataviz (supports visuels) 	<ul style="list-style-type: none"> Les tableaux de bord correspondent à la demande métier : il est opportun de réfléchir à leur mise à jour automatique 	<ul style="list-style-type: none"> Acculturer les métiers aux concepts statistiques
Obstacles	<ul style="list-style-type: none"> Données parfois non disponibles (et difficile de savoir si elles existent) 		<ul style="list-style-type: none"> Risque de doublons ou de confusion (si l'analyse est ponctuelle, c'est Kévin qui récupère les données, si analyse récurrente alors intégration au SI par le Data ingénieur) 	<ul style="list-style-type: none"> Difficulté à implémenter l'automatisation simultanée de la récupération des données + application des traitements de nettoyage 	<ul style="list-style-type: none"> Vulgariser des méthodologies et informations complexes

Retrouvez-nous sur

insee.fr



Marylène Henry

Responsable du domaine Etudes et Synthèses

DSI – Insee

marylene.henry@insee.fr