

CONCEPT NAME

O.P.SÉMIA

SHORT Description

Outil Policier de SÉMantique et d'Investigation Analytique

? Failing to describe your solution? Lost in sterile discussion? Building on too many uncertainties? It's time to go to the Knowledge space!

USAGE SCENARIO

Who are the users of this solution?
How do they actually use it? Describe the key steps of a typical usage scenario.

Utilisation par la police scientifique (SNPS) afin de trier et classer les données issues des scellés.
L'outil se présente comme un moteur de recherche portant sur l'ensemble des données textuelles et visuelles des supports numériques en filtrant et classant les informations par score de pertinence à partir de mots-clés fournis par l'analyste.

KEY WORKING PRINCIPLES

How does this solution work technically?

Après l'extraction depuis les scellés, les données sont filtrées à partir de bases de données répertoriant les annonceurs connus puis les données textuelles sont encodées via un modèle d'embedding local permettant de saisir leur sens général afin de rechercher sémantiquement les informations souhaitées par l'analyste. Un modèle local de vision génèrerait une description de chaque image qui serait ensuite encodée permettant de même sa recherche sémantique.
Lors d'une recherche, les résultats sont présentés dans l'ordre croissant de distance sémantique (score de pertinence entre 100 et -100)

LEARNING TOPICS

Review each aspect of your solution & list here:

- the things you need to investigate to **better define** your solution
- and/or the assumptions you need to **validate** to make sure your solution work effectively.
- Fonctionnement d'un moteur de recherche.
- Modèle d'encodage de texte.
- Modèle de description d'image.
- Renseignements sur les modèles de recherches d'images/reconnaissance faciale
- Modèle NER (Reconnaissance d'entités nommées).
- Evaluation de distance sémantique
- Fonctionnement des systèmes de filtrage d'annonceurs

VALUE proposition

Who benefit from this solution?
What kind of value does this solution bring to them? (users, company, partners, ...)

Le SNPS bénéficie de cette solution en ayant la possibilité de rechercher rapidement et efficacement des champs sémantiques utiles parmi des sources sélectionnées sans que l'analyste n'aie à analyser la totalité des données.
Les données impertinentes ne se seront pas exclues de l'étude mais seulement mises en retrait afin de permettre la relecture et d'éviter les faux négatifs.
Amélioration de la précision sur le score de pertinence grâce à la mise en place de modèles indépendants pour chaque type ou emplacement de fichiers par rapport à un modèle global.

BUSINESS MODEL

How does this solution generate money? (direct sales, subscription, free, advertising revenues, cost reduction...)

Cet outil sera mis a disposition du SNPS gratuitement et permettra de générer un gain d'argent à travers le gain de temps amené par l'augmentation de la productivité de la cellule d'analystes du SNPS.
Plus d'affaires pourront être analysées à temps égal alors que la charge de travail sur les agents diminuera.

FIRST LEARNING ACTIONS

Identify the most critical topic to address:

- first the things you need to learn to be able to describe your solution
 - then the riskiest assumptions, that would disqualify your solution if wrong.
- Formulate it as precisely as possible to optimize your learning effort.

- Fonctionnement d'un moteur de recherche
- Evaluation de distance sémantique

Now be smart!
How can you learn it quickly and at lower cost? And even more quickly?