Activités de recherche en traitement automatique des langues naturelles chez Orange

Atelier TextMine 2019

orange™

Orange en un clin d'œil

1 métier

Opérateur multi-services

5 activités

- Services de communication pour les particuliers
- Services de communication pour les entreprises
- Contenus
- Services financiers mobiles
- Vente en gros

3 familles de clients

- Particuliers
- Entreprises
- Opérateurs et fournisseurs de services

273 millions de clients

28 pays pour les services grand public, et une présence mondiale avec Orange Business Services

152 000 collaborateurs

51e marque mondiale
selon le classement Brand Finance « Global 500 » 2018

22 millions de fans

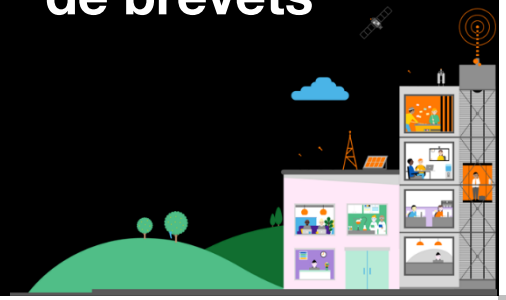
3 millions de followers

1,5 million de visiteurs sur Orange.com chaque mois

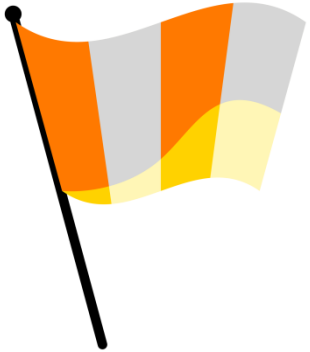
360 661 abonnés

La recherche en quelques chiffres

1er opérateur européen pour les dépôts de brevets



200 brevets déposés en 2017



600 chercheurs

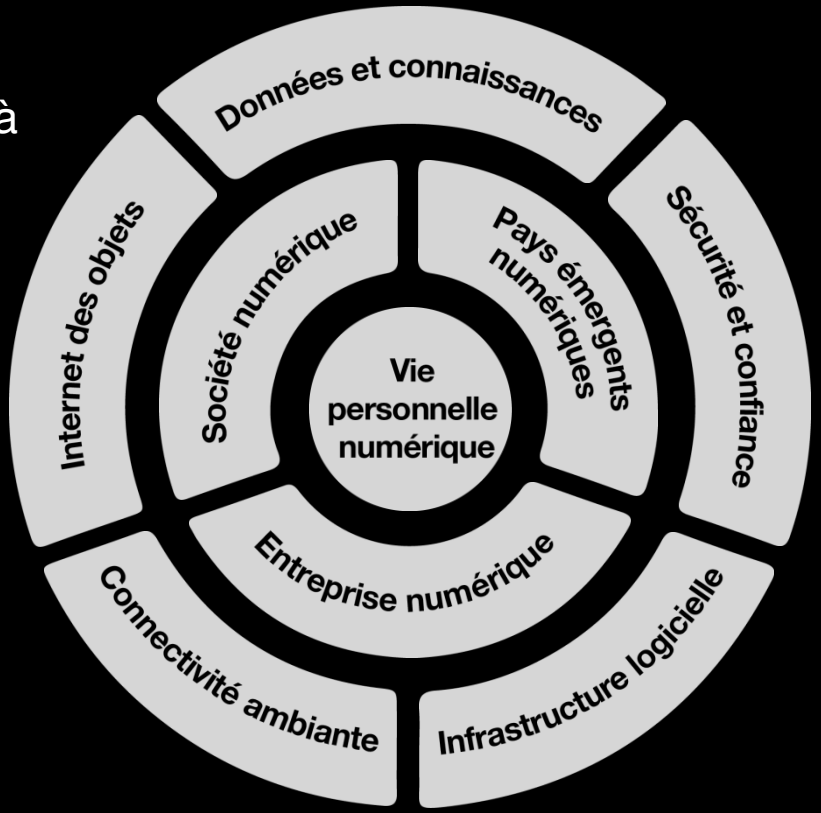
140 doctorants / post-doctorants

100 partenariats (laboratoires, projets coopératifs, IRTs…)

9 domaines de recherche

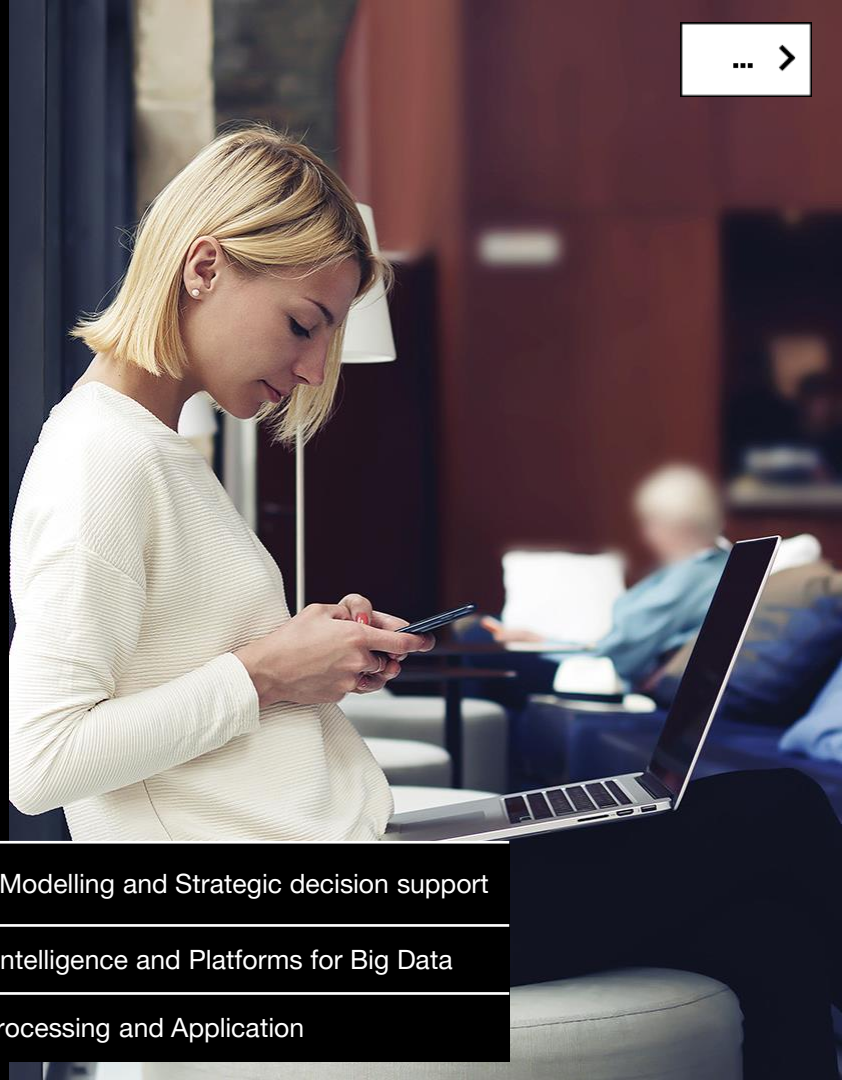
- soutenir la stratégie **Essentiels2020** et au-delà
- piloter nos investissements de recherche
- explorer les ruptures de technologies, d'usages et de modèles économiques

  - via des plateformes de recherche ouvertes et modulaires intégrant technologies et usages à travers des expériences du futur
  - en impliquant dans une recherche à 360° des compétences techno, SI, UX, tech-éco, juridique, marketing…



Données et connaissance

Développer des algorithmes, des briques techniques et des plates-formes respectueuses pour le Big Data, et pour l'intelligence artificielle



Immersive Communication, Capture and Entertainment	Tools for Economic Modelling and Strategic decision support
Recognition in Audio-visual Content	Data management, Intelligence and Platforms for Big Data
Operations Research and Data Analytics	Natural Language Processing and Application

20 ans de recherche en TALN chez Orange

- évolution des technologies
- évolution des métiers
- évolution des activités
- des textes de plus en plus variés

Répondre aux besoins en TALN du groupe

- français
- anglais
- espagnol
- polonais
- néerlandais
- arabe
- roumain

- corriger
- extraire
- classer
- comprendre

- documents pour la communication
- documents pour la production
- requêtes
- commentaires
- verbatims
- posts/tweets
- chats

Les activités de recherche en TALN chez Orange

Eclairer le futur, identifier les risques et les ruptures, explorer et construire des opportunités, influencer nos écosystèmes

veille scientifique

communication

éco système TALN

Bâtir des actifs stratégiques, compétences, propriété intellectuelle, standards, partenariats

logiciels

brevets

ressources

partenariats

Alimenter l'innovation pour nos produits et services, mais aussi pour améliorer l'efficacité de nos infrastructures

expertise pour les projets

solutions sur mesure

solutions génériques

Organisation par projets de recherche

Recherche textes

LIS — CRE, CET

IRIT — CRE

Intégration

FMSH fondation maison des sciences de l'homme

Recherche dialogue

LIA LABORATOIRE INFORMATIQUE D'AVIGNON — CRE

Loria Laboratoire lorrain de recherche en informatique et ses applications — CET

ANR PASTEL

lium

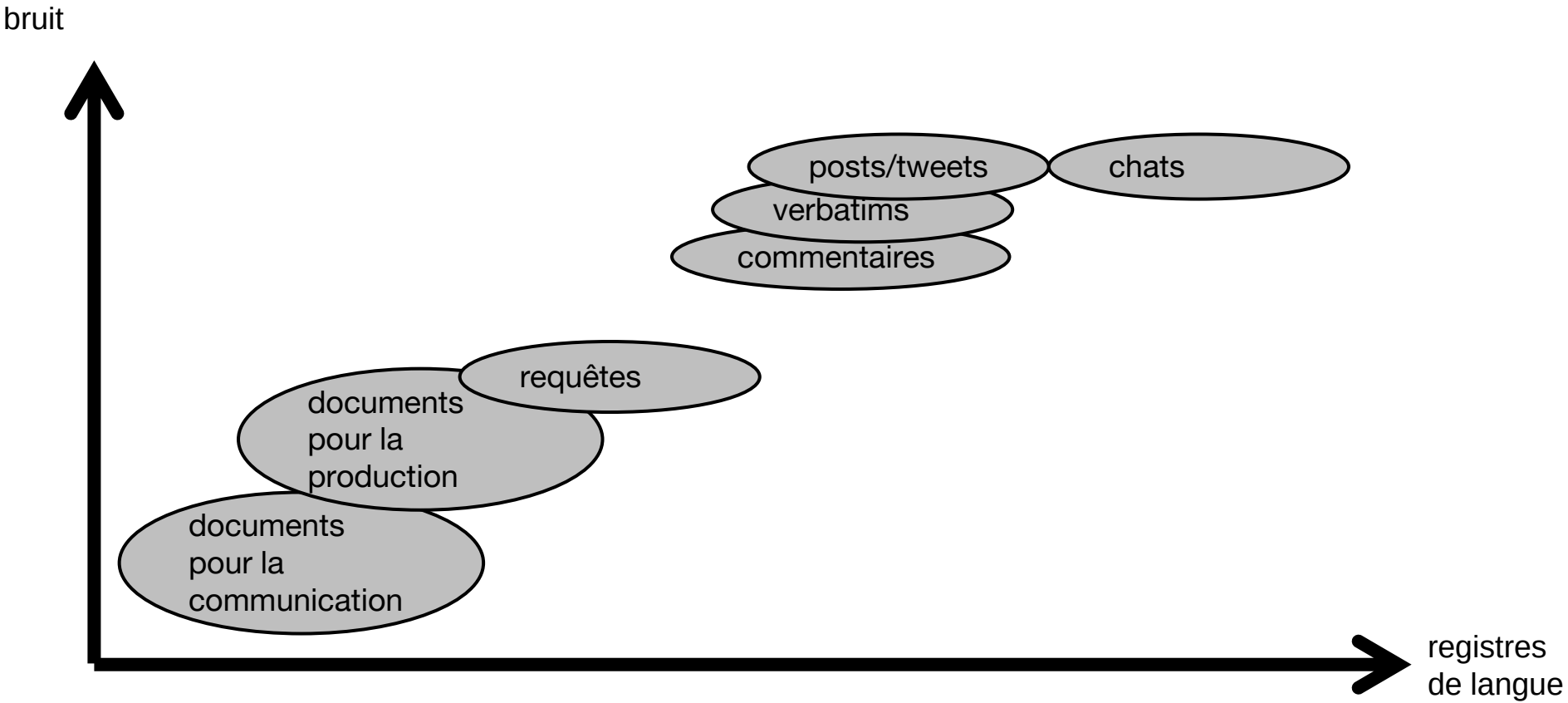
LS2N

ANR DATCHA

IRIT

LIS LABORATOIRE D'INFORMATIQUE & SYSTÈMES

Les types de textes chez Orange



Exemples de besoins exprimés par type de texte

	corriger	extraire	classer	comprendre
Chats	correction	mots-clés, anonymisation	Polarité, Labels, Clustering	
Posts/tweets				
Verbatim				
Commentaires				
Requêtes		entités	type réponses, similarité	relations
Documents pour la production				
Documents pour la communication				

Quelques sujets de recherche

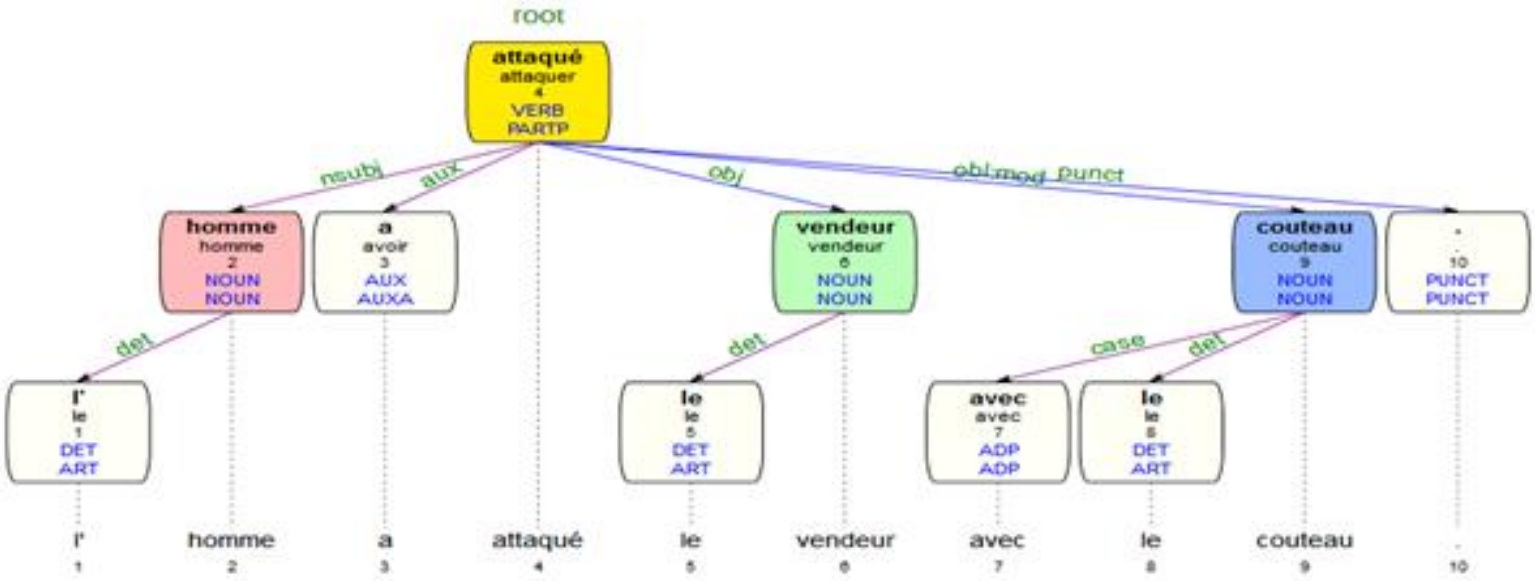
requêtes

documents pour la production

documents pour la communication

Analyse en dépendance

- l' homme a attaqué le vendeur avec le couteau .



CoNLL 2017 Shared Task on Multilingual Dependency Parsing : 10th out of 33

Analyse en rôles sémantiques

requêtes

documents pour la production

documents pour la communication

- l' homme a attaqué le vendeur avec le couteau .

Relation(s)

- Attack(Assaillant homme, Victim vendeur, Weapon couteau)

Quadruple(s)

graph	subj	pred	obj
g4	Attack	Assaillant	homme
g4	Attack	Victim	vendeur
g4	Attack	Weapon	couteau

verbe agent (celui qui fait) patient (celui qui subit) frame target ou frame element

	La Vénus de Milo	a été	transportée	rapidement	en bateau	de la Turquie	au musée du Louvre	par Jules Dument	en 1821
Syntaxe	nsubj:pass		root	advmod	obl:en	obl:de	obl:au	obl:par	obl:en
SRL	Patient		target	Manner	Instrument	Ablative	Location	Agent	Time
PropBank	Arg1		LU		Arg4	Arg3	Arg2	Arg0	Arg-TMP
Frames	Theme		Frame:Motion	Manner	Means	Source	Goal	Carrier	Time

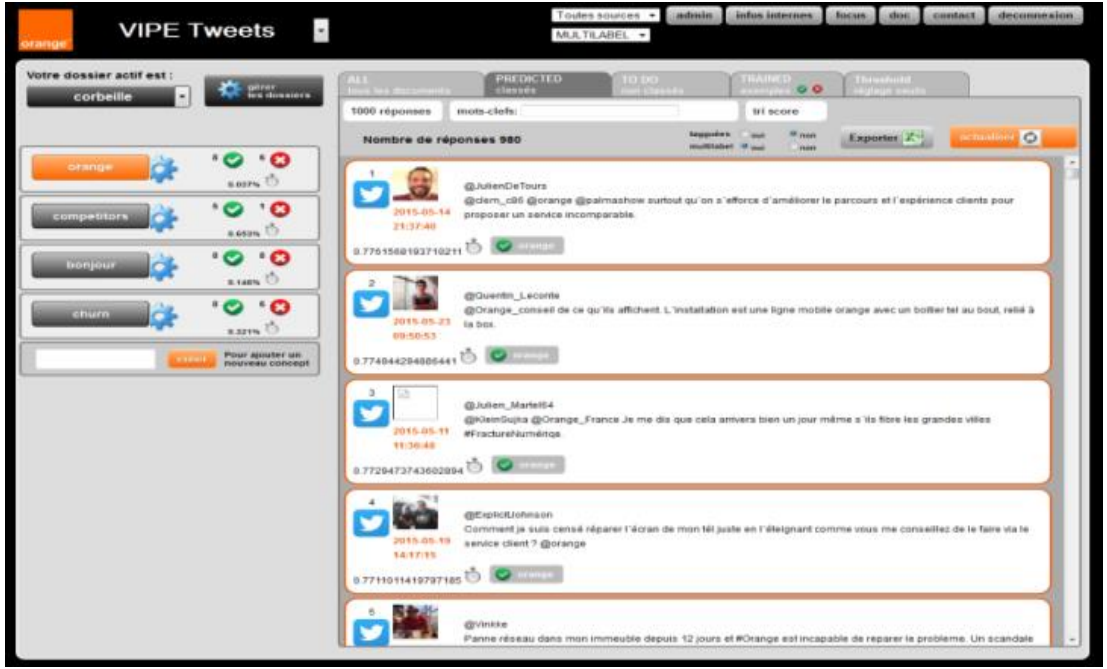
relations

ViPE : labélisation de textes courts

commentaires

verbatims

posts/tweets



VIPE : un outil interactif de classification multilabel de messages courts
- Frank Meyer, Sylvie Tricot, Pascale Kuntz, Wissam Siblini – EGC 2017

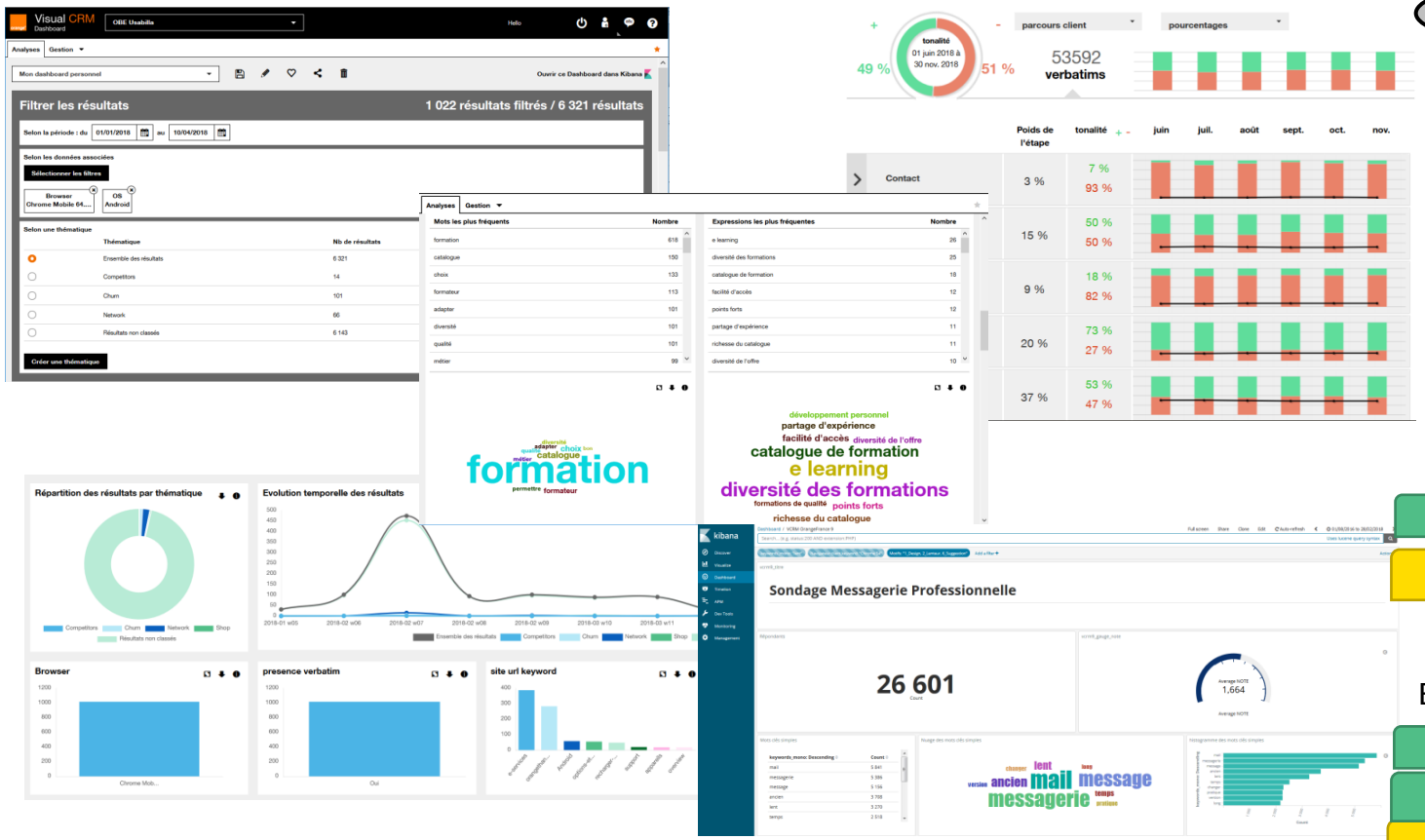
Labels

VisualCRM

commentaires

verbatims

posts/tweets



Labels

mots-clés

Besoins :

Clustering

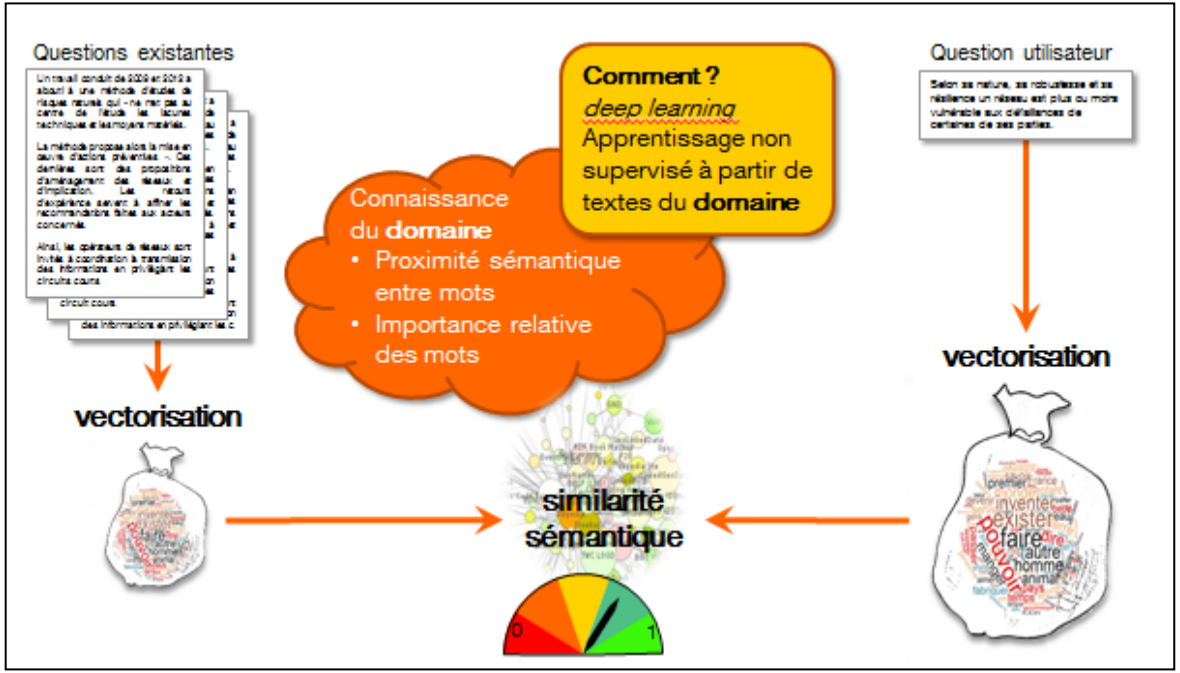
similarité

anonymisation

requêtes

Simbow : similarité entre textes courts

de la proximité entre mots à la proximité entre textes ou comment mesurer une similarité entre textes même s'ils n'ont aucun mot en commun?



International Challenge SemEval 2017: 1st on Community Question Answering (Task 3, subtask B)

similarité

requêtes

Ask’O : questions/réponses

Ask'O

En électricité, quelle est l'unité de mesure de la fréquence ?

Compagnon

hertz

Quelle est la langue officielle de l'Angola ?

Compagnon

portugais

Qui est la mère de l'actrice Lou Doillon ?

Compagnon

la mère de l' actrice Lou Doillon est Jane Birkin

relations

Explorescence

transcription

documents pour la communication

conférence video en français du 13 novembre 2014 d'une durée de 29m
collection : Institut national de recherche en informatique et en automatique / Enjeux du web 3.0 dans le s...
sciences humaines sociales

Faire face aux enjeux politiques et technologiques avec le web de données. Ouverture
Quel avenir pour le Web ?

Locutions	Locutions étendues	Mots simples	Personnes

- ministère de la culture — 11 occurrences
- web de données — 8 occurrences
- web sémantique — 4 occurrences
  - le web sémantique dans le domaine de la de la culture donc qui a eu un rôle tout à fait important — 00:05:14
  - du web des données liées et du web sémantique il était tout à fait — 00:09:05
  - qui vont être des données liées et du web sémantique — 00:09:34
  - à travers une feuille de route stratégique métadonnées culturel et transition web trois virgule 0 précise que ce qu'on entend par trois virgule 0 n'est rien d'autre que le web de données liées web sémantique qui vise à définir des axes thématiques pour mettre en place des projets et parmi ces accidents n'a Gandon Fabien on va vous parler dans quelques instants — 00:10:49

tri par fréquence

Détection des locutions
à partir de la transcription automatique
« web sémantique »

00:10:52
00:28:46

http://pits-fmsh.orange-labs.fr/canalu/collection

mots-clés

Publications 2018

- ICML (International Conference on Machine learning)
  - CRAFTML, an Efficient Clustering-based Random Forest for Extreme Multi-label Learning
    Wissam Siblini, Pascale Kuntz & Frank Meyer
- CORIA-TALN
  - Similarité textuelle pour l'association de documents journalistiques
    Delphine Charlet et Géraldine Damnati
  - Evaluation automatique de la satisfaction client à partir de conversations de type « chat » par réseaux de neurones récurrents avec mécanisme d'attention
    Jeremy Auguste, Delphine Charlet, Géraldine Damnati, Benoit Favre et Frederic Bechet
  - Predicting failure of a mediated conversation in the context of asymetric role dialogues [POSTER]
    Romain Carbou, Delphine Charlet, Géraldine Damnati, Frédéric Landragin and Jean Léon Bouraoui
- LREC (Language Resources and Evaluation Conference)
  - Handling Normalization Issues for Part-of-Speech Tagging of Online Conversational Text
    Géraldine Damnati, Jérémy Auguste, Alexis Nasr, Delphine Charlet, Johannes Heinecke and Frédéric Béchet
  - Semantic Frame Parsing for Information Extraction : the CALOR corpus
    Gabriel Marzinotto, Jeremy Auguste, Frederic Bechet, Geraldine Damnati and Alexis Nasr
  - FrNewsLink : a corpus linking TV Broadcast News Segments and Press Articles
    Nathalie Camelin, Géraldine Damnati, Abdessalam Bouchekif, Anais Landeau, Delphine Charlet and Yannick Esteve
- SLSP 2018 (Statistical Language and Speech Processing)
  - Safe transfer learning for dialogue application
    Nicolas Carrara, Romain Laroche, Jean-Léon Bouraoui, Tanguy Urvoy, Olivier Pietquin

https://orange.jobs

47 offre(s) | créer une alerte

map | Innovation, R&D | CDI | Thèse | Post D

Ingénieur ingénieure de recherche en systèmes de gestion de bases de données graphes réparties et temporelles pour l'Internet des Objets F/H
France - Cesson Sévigné - Orange

Architecte de solutions d'Intelligence Artificielle - F/H
France - Lannion - Orange

Architecte de solutions d'Intelligence Artificielle - F/H
France - Chatillon - Orange

Ingénieur de Recherche en Intelligence Artificielle pour l'étude de la mobilité dans les réseaux mobiles F/H
France - Belfort - Orange

Ref : 0019513 | 02 janv. 2019

Date limite de candidature : 31 janv. 2019

2 avenue Pierre Marzin 22300 LANNION - France

Votre rôle

Vous rejoignez le centre de compétences en Intelligence artificielle au sein du département “Data Intelligence et Algorithmes” d'Orange Labs Services.

Cette équipe composée d'experts IA avec un réel savoir-faire algorithmique et opérationnel a pour mission :

- d'aider nos directions métiers et nos filiales à développer l'usage de l'intelligence artificielle et du big data dans leur métier, soit lors de leur phase d'appropriation soit dans la phase de mise en oeuvre d'applications requérant une

Merci