

VIPE

CNAM – 6 mars 2015

Frank Meyer

Orange Labs / IMT / UCE / CRM-DA / PROF



VIPE

The screenshot displays the VIPE (Visual, Interactive and Personalized Exploration of content catalogs) interface. At the top, it shows the Orange logo, the title "VIPE VIPE twitter version 1.3 (Visual, Interactive and Personalized Exploration of content catalogs)", and the date/time "Lundi 12 Mai 2014 11:08:20" along with the user email "franck.meyer@orange.com". Navigation buttons include "actualiser", "mots-clefs", "rechercher", and "déconnexion".

Les concepts franck

- Churn-related: 7 likes, 2 dislikes, 0 shares
- Orange TM: 5 likes, 8 dislikes, 0 shares
- Négatif, insultant: 4 likes, 3 dislikes, 0 shares
- Concurrence**: 4 likes, 6 dislikes, 0 shares
- concurrence insultant: 0 likes, 2 dislikes, 0 shares
- en anglais: 2 likes, 2 dislikes, 0 shares
- problèmes freenauts: 11 likes, 32 dislikes, 0 shares
- créer

Les tweets proches de "Concurrence"

- Juventin DROMER** (Saint-Dizier, France) - 14/04/2014 15:26:13 [0.700]: Rachat de SFR : Martin Bouygues estime avoir été floué par Vivendi <http://t.co/N703Bjzht> #pcinpact
- [Profile]** (La Rochelle) - 02/04/2014 18:13:04 [0.663]: @patrick_ca_ma @maxoubx @sosh_fr @infestedgrunt c'est confirmé par Orange bug de remise à zero ;) Free ne peut pas répondre à ça #troll
- [Profile]** - 22/04/2014 06:05:00 [0.630]: RT @Boobiche: @boom3120 SFR mais tkt j'ai fais Orange et Free aussi et c'est pareille, c'est l'endroit ou tu vis qui compte
- [Profile]** (Montpellier) - 13/04/2014 18:50:32 [0.627]: @MlleLiLaa @rousseau__ moi c'est sur la boite mail principale orange celle de mon père
- [Profile]** (Rouen) - 26/02/2014 10:58:51 [0.615]: @sfr incompétence totale, gestion calamiteuse des dossiers, SAV catastrophique... C'est très grave!!!!
- [Profile]** (Le Mans) - 05/05/2014 19:52:03 [0.607]: Quel opérateur choisir pour ma box, SFR ? FREE ? ORANGE ?
- Guilherand-** - 28/03/2014 12:02:14 [0.590]: @Alexzander_F @LALIGNEDEFREE Chez bouygues ça marche très bien, et la couverture 4G est beaucoup plus importante que celle de Free (75%);





- Application prototype pour l'apprentissage multi-label interactif
 - 2 sous-applications en ligne (utilisable par vous pour des tests)
 - **Vipe-VOD** : démonstrateur de système de recommandation multi-profil ou multicritères
 - **Vipe-Tweet** : démonstrateur de système de reconnaissance de concepts sur petits textes très bruités, les tweets
 - Développée en mode application web
 - PHP pour l'applicatif WEB, + HTML5 et Javascript
 - MySQL (pour le moment) pour le SGBD
 - Algorithme d'apprentissage en JAVA
 - Par la suite, on approfondit uniquement Vipe-Tweet

VIPE-Tweet : particularités

VIPE



- Principales particularités de VIPE-Tweet
 - Prototype text-mining temps-réel
 - Apprentissage interactif ↔ apprentissage quasi temps réel
 - Fonctionnement via une IHM intuitive (pas de cycle batch modélisation / application du modèle)
 - Vipe classe tous les tweets dans les catégories apprises (soit comme exemple positif, soit comme exemple négatif) ; les mises à jour sont prises en compte en quasi temps-réel
 - Démonstrateur : système de classification et de veille de la twittosphere
 - surveillance des villes toutes les heures, récupère des tweets autour de ces villes, classe les tweets selon des catégories personnalisées

VIPE-Tweet

recherche de tweets par mots-clés (pour ensuite les donner en exemples ou contre-exemples)

VIPE



Utilisateur / profil actuel

VIPE twitter version 1.3 (Visual, Interactive and Personalized Exploration of content catalogs)

Lundi 12 Mai 2014 11:08:20
franck.meyer@orange.com

actualiser mots-clefs rechercher déconnexion

enregistrer imprimer

Les concepts franck

Churn-related	7	2	0	0
Orange TM	5	8	0	0
Négatif, insultant	4	3	0	0
Concurrence	4	6	0	0
concurrence insultant	0	2	0	0
en anglais	2	2	0	0
problèmes freenautes	11	32	0	0
créer				

Les tweets proches de "Concurrence"

- 14/04/2014 15:26:13 [0.700] Rachat de SFR : Martin Bouygues estime avoir été floué par Vivendi <http://t.co/N703Bjzht> #pcinpact Saint-Dizier, France
- 02/04/2014 18:13:04 [0.663] @patrick_ca_ma @maxoubx @sosh_fr @infestedgrunt c'est confirmé par Orange bug de remise à zero ;) Free ne peut pas répondre à ça #troll La Rochelle
- 22/04/2014 06:05:00 [0.630] RT @Boobiche: @boom3120 SFR mais tkt j'ai fais Orange et Free aussi et c'est pareille, c'est l'endroit ou tu vis qui compte
- 13/04/2014 18:50:32 [0.627] @MlleLiLaa @rousseau__ moi c'est sur la boite mail principale orange celle de mon père Montpellier
- 26/02/2014 10:58:51 [0.615] @sfr incompétence totale, gestion calamiteuse des dossiers, SAV catastrophique... C'est très grave!!!! Rouen
- 05/05/2014 19:52:03 [0.607] Quel opérateur choisir pour ma box, SFR ? FREE ? ORANGE ? Le Mans
- 28/03/2014 12:02:14 [0.590] @Alexzander_F @LALIGNEDEFREE Chez bouygues ça marche très bien, et la couverture 4G est beaucoup plus importante que celle de Free (75% ;) Guilhaudon

sélection courante

Tweets correspondant à la sélection courante (parlant de « concurrence »)

Voyant vert : le concept est appris (les scores prédictifs sont à jour)

Nombre d'exemples positifs déjà donnés pour « problèmes freenautes »)

Nombre d'exemples négatifs donnés pour « problèmes freenautes »)

VIPE-Tweet

VIPE



Cliquer sur un tweet pour l'ouvrir...

Les concepts franck

- Churn-related 7 ✓ 2 ✗
- Orange TM 5 ✓ 8 ✗
- Négatif, insultant 4 ✓ 3 ✗
- Concurrence 4 ✓ 6 ✗
- concurrence insultant 0 ✓ 2 ✗
- en anglais 2 ✓ 1 ✗
- problèmes freenautes 11 ✓ 32 ✗

créer

L'actualité des tweets avec recherche de "orange"

enregistrer imprimer

Nantes

BX 06/05/2014 17:27:45 @Orange_France Venez fibrer ma maison svp bisous ? Cliquer pour ouvrir le tweet

Grenoble

06/05/2014 Trierweiler "souhaite bonne chance" à François Hollande - #actu <http://t.co/3TNr211gXv> via @Vidéo avec Dailymotion d'Orange

Metz

06/05/2014 Go Behind Bars With The Craziest Inmate On 'Orange Is The New Black' <http://t.co/O8s4K4L83g>

Paris France

12/05/2014 Keyshop today : Bonnet tressé à 2 pompons Bonnet tressé à 2 pompons, 16:10:44 disponible en rouge, noir, blanc et orange... <http://t.co/OU6qrxg2Ls>

Paris

12/05/2014 #1: Burt's Bees Orange Essence Facial Cleanser, 4.3 Ounces <http://t.co/2t173SU7M6>Burt's Bees Orange Essence F... <http://t.co/49KiUVRHoz>

us

10/05/2014 RT @GozziPaul: les meufs elles mettent 2 tonnes de maquillages elle sont orange 09:57:50 du visage et blanche du cou on dirait des surimi

08/05/2014 RT @CNET: Hackers nab data on 1.3M Orange telco customers <http://t.co/lc7CuQi6neR>

03:11:13

focus paramètres aide nous écrire

VIPE-Tweet

... puis classer le tweet, positivement ou négativement, dans les catégories souhaitées

VIPE



Les concepts franck

Churn-related	7	2	
Orange TM	6	8	
Négatif, insultant	4	4	
Concurrence	4	6	
concurrence insultant	0	2	
en anglais	2	1	
problèmes freenautes	11	32	
créer			

L'actualité des tweets avec recherche de "orange"

enregistrer imprimer

Fermer la fenêtre

description: Je me sent tellement bien dans la noirceur [me sens dans mon élément] @Stromae @linkinpark, @Muse
texte: @Orange_France Venez fibrer ma maison svp bisous ?

Churn-related 0.06	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Orange TM 1.00	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Négatif, insultant 0	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Concurrence 0.18	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
concurrence insultant 0.80	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
en anglais 0.16	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
problèmes freenautes 0.11	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

ordonner par score

focus paramètres aide nous écrire

VIPE-Tweet

FOCUS : spécifier les mots clés et les lieux complémentaires qui doivent être suivis

VIPE



Les concepts franck

- Churn-related 7 2
- Orange TM 6 8
- Négatif, insultant 4 4
- Concurrence 4 6
- concurrence insultant 0 2
- en anglais 2 1
- problèmes freenautes 11 32

Focus : entrez des mots-clefs ou des lieux

Ces mots ou lieux correspondants à vos centres d'intérêt enrichiront la recherche des nouveaux tweets.

- neural
- decision tree
- mining
- machine learning
- dreamtime
- uluru
- bird

Les données doivent être décimales. Pour convertir de degrés en décimales, allez sur le site : <http://www.care-is.com/fr/12/outils/17/convertisseur-gps.html>

Nom du lieu	Latitude	Longitude	
Uluru	-25.3451	131.0364	<input type="button" value="supprimer"/>
Boston	42.3577	-71.061	<input type="button" value="supprimer"/>
Kiev	50.4536	30.5163	<input type="button" value="supprimer"/>

problèmes freenautes 0.11

ordonner par score

[focus](#) [paramètres](#) [aide](#) [nous écrire](#)



VIPE-Tweet : à l'intérieur

VIPE



- VIPE est une application WEB composée de 5 modules
 - Module de récupération des Tweets
 - Toutes les heures, va chercher, sur + de 100 villes de France (plus San Francisco, Londres et Madrid) des Tweets
 - ... au hasard, ... ou correspondant à des « focus » utilisateurs
 - Module de gestion de la base de données
 - Intègre les Tweets, gère la matrice Tweet X Mots-clés, efface les anciens Tweets ; gère les utilisateurs et leurs concepts (exemples et contre-exemples)
 - Module Applicatif Web
 - Gère l'IHM
 - Module de scoring
 - Pré calcule les scores prédictifs pour les concepts, mise à jour toutes les 30s
 - Algorithme prédictif de type anytime (Molecule)
 - non détaillé ici.

VIPE : principe de l'apprentissage

VIPE



- Représentation des Tweets et des concepts
 - Grande matrice creuse
 - 2 zones : descriptions des Tweets (données explicatives, logs des concepts (données cibles))
 - apprentissage fait par cycles, proportionnellement 1/10 du temps sur les mots clés, 9/10 sur les concepts

	Mot clé n°1	Mot-clé n°2	Orange	SFR	Jus	Shosh	Free	Problème	couleur	Mairie	passer à	cher	Mot-clé N	Concept 1	Concept 2	Orange TM	Concurrence	Churn (changement opérateur)	Concept N
Tweet n°1	1	1						1								1	1		
Tweet n°2				1							1	1						1	1
			1			1			1								1		
				1													1		
Tweet n°i				1				1		1	1	1	1	1	1				
	1				1			1											
																1	1	1	
Tweet n° M			1			1			1										1

On utilise un algorithme de factorisation rapide de matrice pour apprendre en quasi-temps réel.

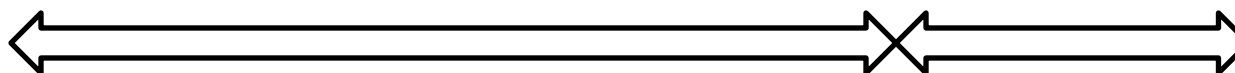


- Volumétrie (mai 2014)

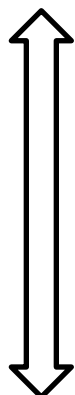
une passe d'apprentissage dure environ 15 secondes

environ 222 000 mots clés (colonnes)

quelques dizaines de concepts



environ
167 000
tweets
(lignes)



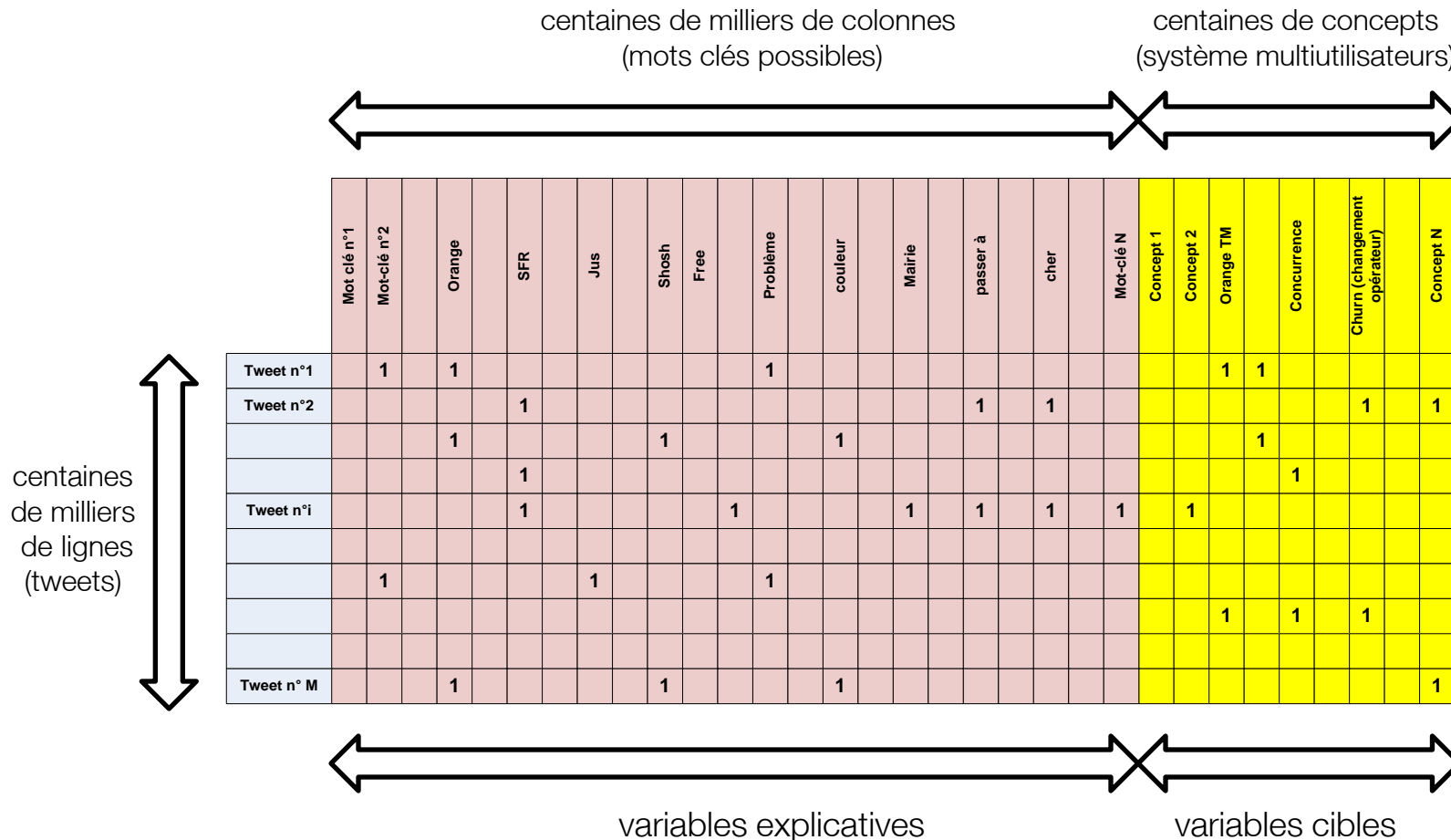
	Mot clé n°1	Mot-clé n°2	Orange	SFR	Jus	Shosh	Free	Problème	couleur	Mairie	passer à	cher	Mot-clé N	Concept 1	Concept 2	Orange TM	Concurrence	Churn (changement opérateur)	Concept N	
Tweet n°1	1	1						1								1	1			
Tweet n°2				1							1	1						1		1
			1			1			1								1			
Tweet n°i				1				1		1	1	1	1	1						
	1				1			1												
																1	1	1		
Tweet n° M			1				1		1											1

environ 1 400 000 logs liés aux contenus des tweets
(moyenne de 8 mots clés par tweets environ)

environ 800 logs liés aux
exemples / contre-exemples
de concepts utilisateurs



- VIPE est un bon cas d'usage pour expérimenter des systèmes interactifs multi-labels (multi-cibles) sur données massives



POUR TESTER VIPE :

<http://prof.orange-labs.fr/vipe/index.php>

