SÉRENDIPITÉ DU MAUVAIS ÉLÈVE

Frédéric Clavert - frederic.clavert@unil.ch

LA RECHERCHE #WW1: TRACES DU CENTENAIRE SUR TWITTER

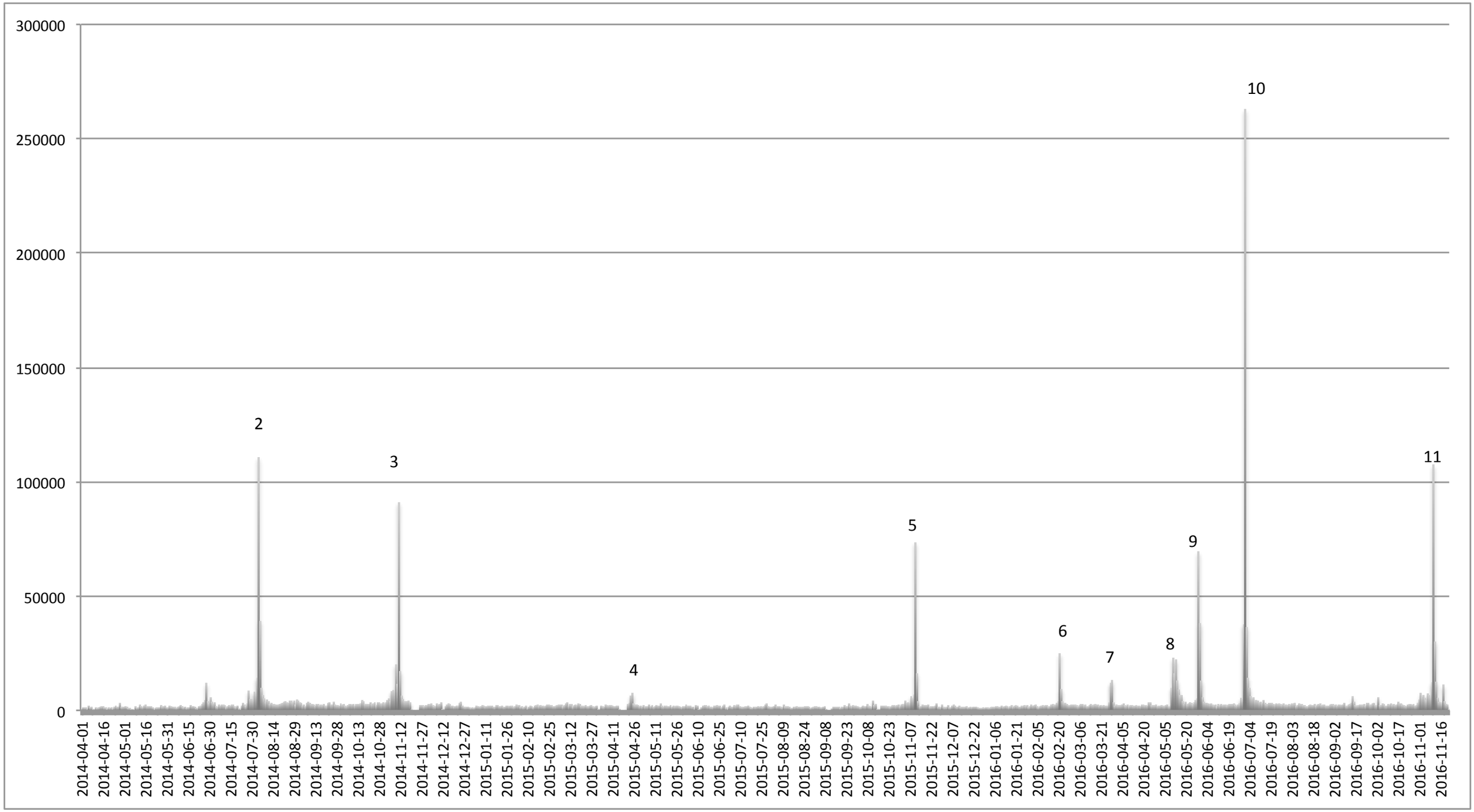
- Traces de mémoire collective sur les réseaux sociaux numériques, notamment Twitter
- Collecte de tweets liés au Centenaire de la Grande Guerre ou à la Grande Guerre elle-même
  - Commencée le 1er avril 2014
  - Dans l'idéal, se terminera fin juin 2019 (centenaire du traité de Versailles)
  - Sur la base de mots-clés en anglais, français, allemand
- Techniquement: un script se connecte à l'API dite de *streaming* de Twitter
  - Jusqu'à 1% du flux des tweets (500 millions / jours)
  - Limite jamais atteinte pour le projet #ww1

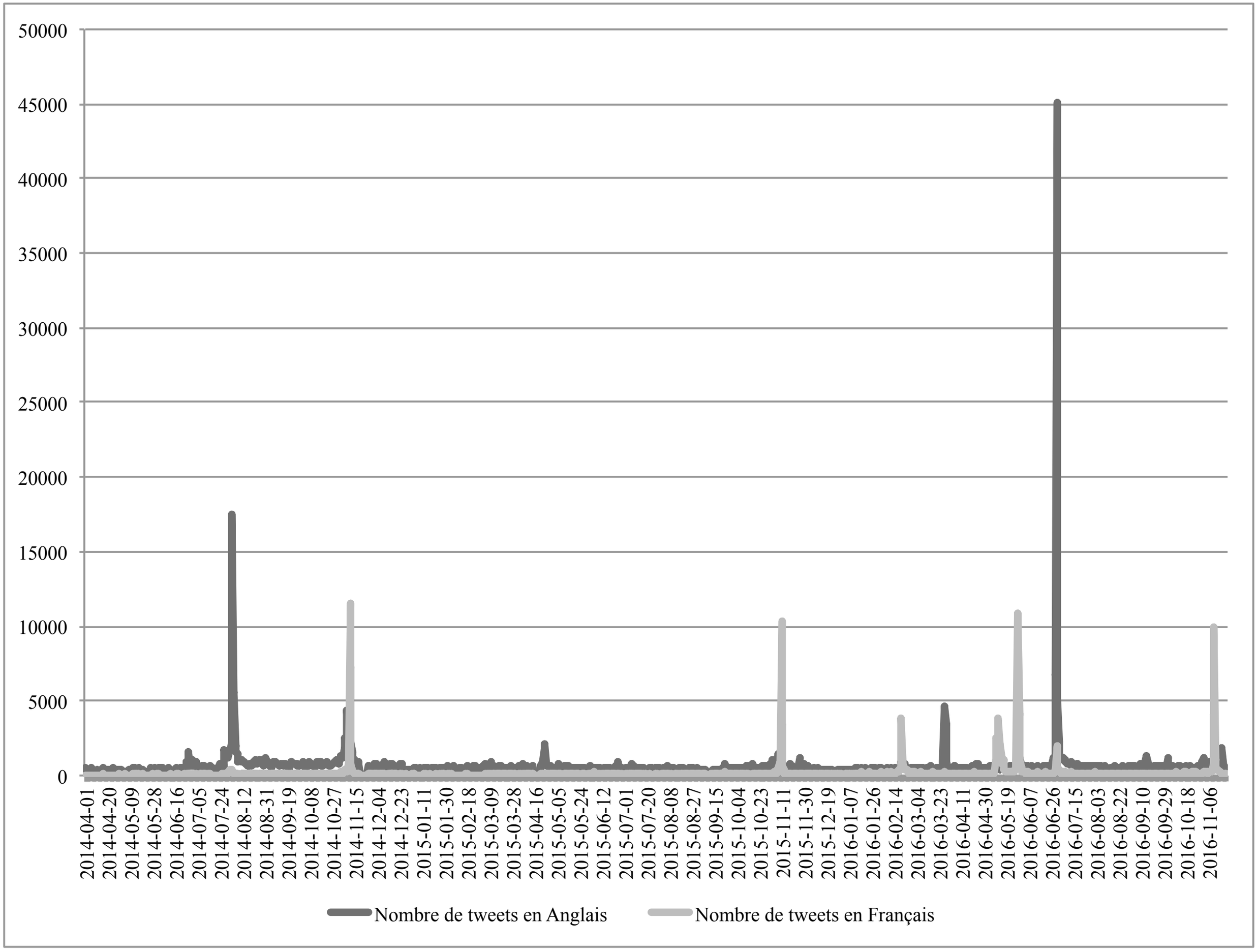
ETAT DE LA BASE DE DONNÉES (NOVEMBRE 2016)

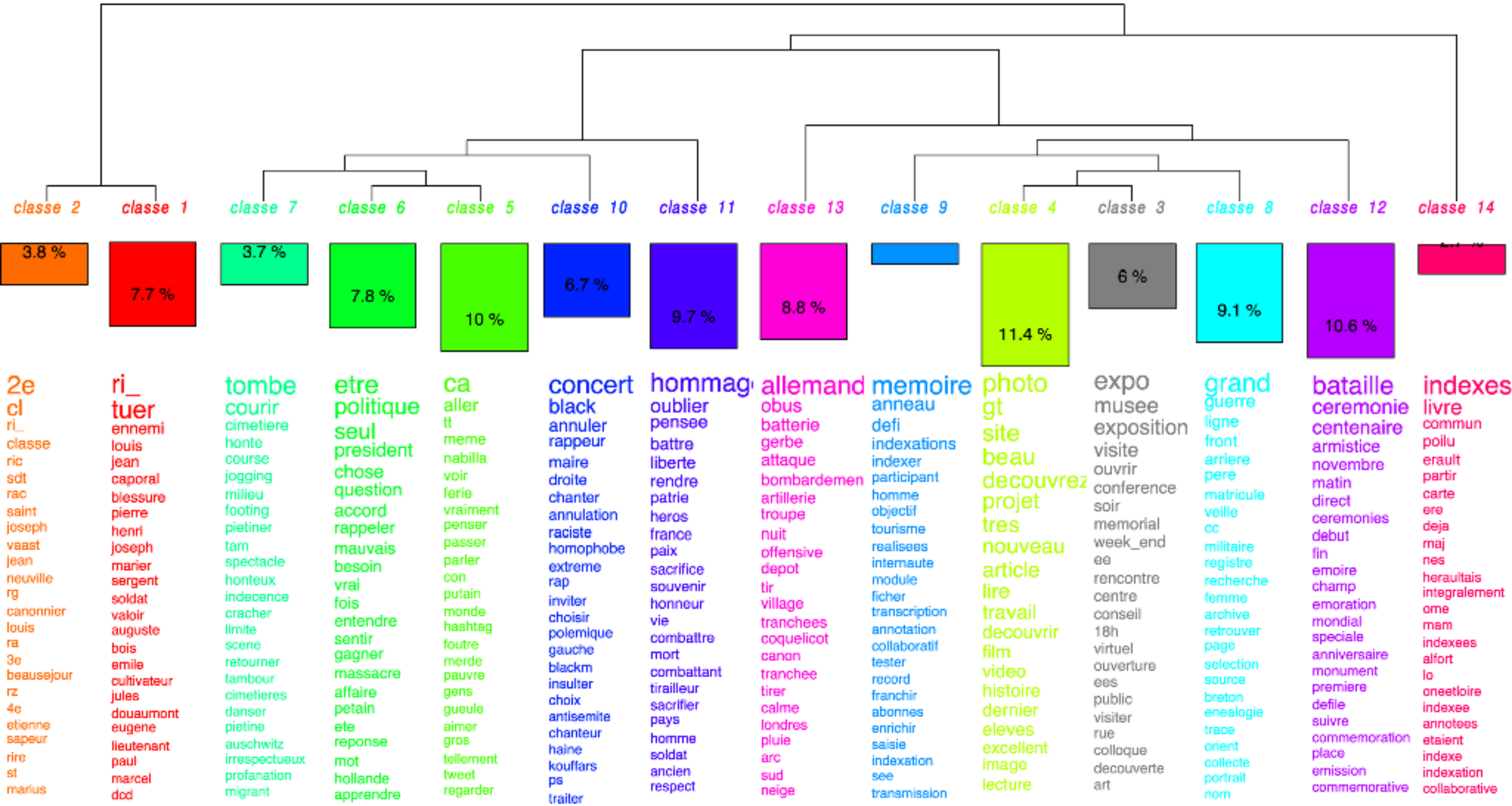
- 3,5 millions de tweets
  - 1,1 million de tweets «originaux»
  - 2,4 de retweets
- Un peu plus de 840 000 utilisateurs
  - Données publiques
- 140 000 hashtags différents
  - 61 421 n’apparaissent qu’une seule fois
  - 120 000 apparaissent dix fois ou moins.

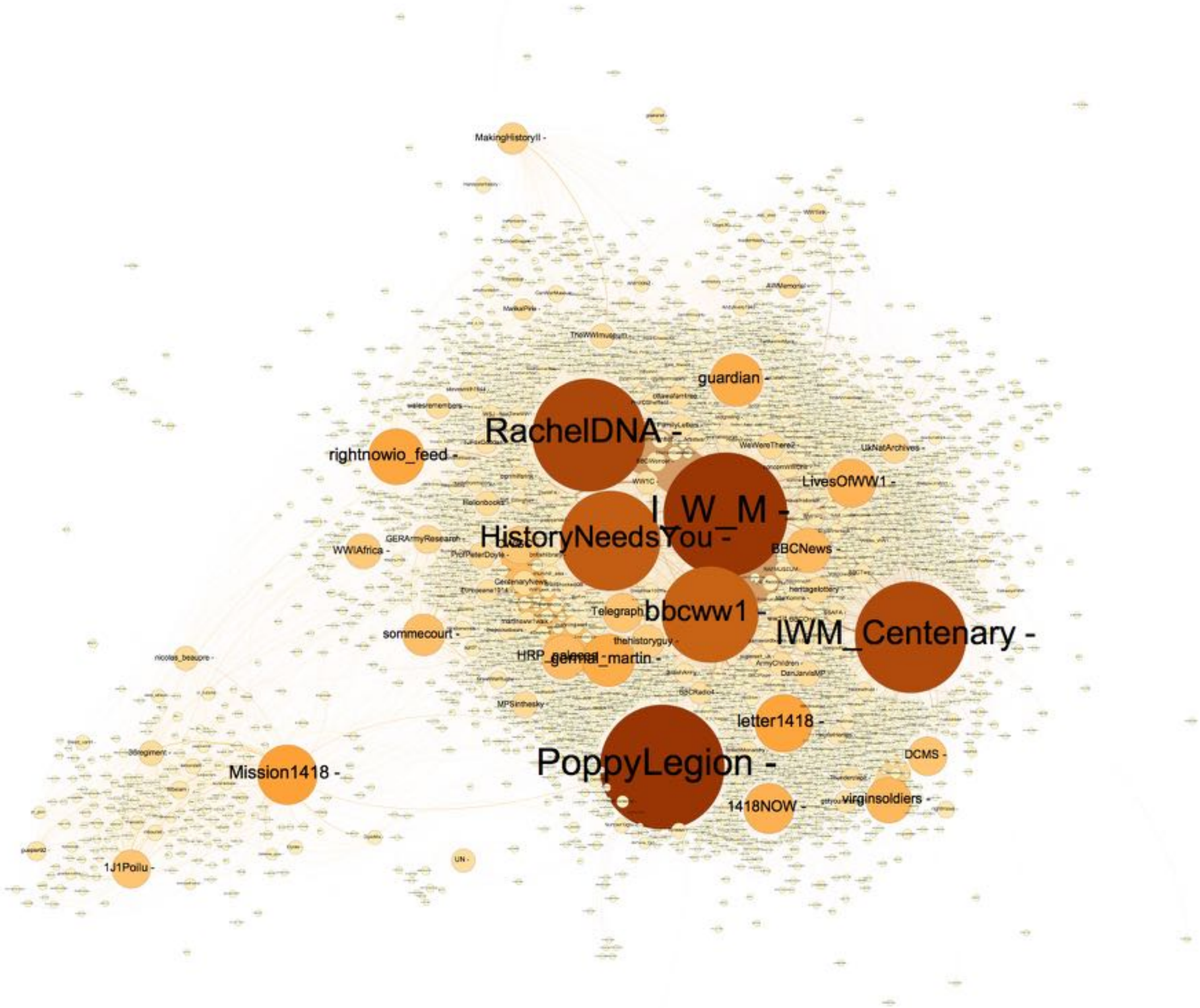
QUELQUES RÉSULTATS

Mot-clé (hashtag)	Nombre de tweets
ww1	1636319
somme100	366827
wwi	257203
11novembre	184970
verdun	173582
1gm	126145
somme	122771
ww1centenary	120527
poppies	120225
11nov	105954
verdun2016	96167
fww	74223
lestweforget	69507
asmsg	65733
1j1p	62952
history	61374
onthisday	50834
centenary	50615
100years	45433
blackm	42509


300000
250000
200000
150000
100000
50000
0
2
3
4
5
6
7
8
9
10
11
2014-04-01
2014-04-16
2014-05-01
2014-05-16
2014-05-31
2014-06-15
2014-06-30
2014-07-15
2014-07-30
2014-08-14
2014-08-29
2014-09-13
2014-09-28
2014-10-13
2014-10-28
2014-11-12
2014-11-27
2014-12-12
2014-12-27
2015-01-11
2015-01-26
2015-02-10
2015-02-25
2015-03-12
2015-03-27
2015-04-11
2015-04-26
2015-05-11
2015-05-26
2015-06-10
2015-06-25
2015-07-10
2015-07-25
2015-08-09
2015-08-24
2015-09-08
2015-09-23
2015-10-08
2015-10-23
2015-11-07
2015-11-22
2015-12-07
2015-12-22
2016-01-06
2016-01-21
2016-02-05
2016-02-20
2016-03-06
2016-03-21
2016-04-05
2016-04-20
2016-05-05
2016-05-20
2016-06-04
2016-06-19
2016-07-04
2016-07-19
2016-08-03
2016-08-18
2016-09-02
2016-09-17
2016-10-02
2016-10-17
2016-11-01
2016-11-16


50000
45000
40000
35000
30000
25000
20000
15000
10000
5000
0
2014-04-01
2014-04-20
2014-05-09
2014-05-28
2014-06-16
2014-07-05
2014-07-24
2014-08-12
2014-08-31
2014-09-19
2014-10-08
2014-10-27
2014-11-15
2014-12-04
2014-12-23
2015-01-11
2015-01-30
2015-02-18
2015-03-09
2015-03-28
2015-04-16
2015-05-05
2015-05-24
2015-06-12
2015-07-01
2015-07-20
2015-08-08
2015-08-27
2015-09-15
2015-10-04
2015-10-23
2015-11-11
2015-11-30
2015-12-19
2016-01-07
2016-01-26
2016-02-14
2016-03-04
2016-03-23
2016-04-11
2016-04-30
2016-05-19
2016-06-07
2016-06-26
2016-07-15
2016-08-03
2016-08-22
2016-09-10
2016-09-29
2016-10-18
2016-11-06
Nombre de tweets en Anglais
Nombre de tweets en Français


classe 2
3.8 %
2e
cl
ri_
classe
ric
sdt
rac
saint
joseph
vaast
jean
neuville
rg
canonnier
louis
ra
3e
beausejour
rz
4e
etienne
sapeur
rire
st
marius
classe 1
7.7 %
ri_
tuer
ennemi
louis
jean
caporal
blessure
pierre
henri
joseph
marier
sergent
soldat
valoir
auguste
bois
emile
cultivateur
jules
douaumont
eugene
lieutenant
paul
marcel
dcd
classe 7
3.7 %
tombe
courir
cimetiere
honte
course
jogging
milieu
footing
pietiner
tam
spectacle
honteux
indecence
cracher
limite
scene
retourner
tambour
cimetieres
danser
pietine
auschwitz
irrespectueux
profanation
migrant
classe 6
7.8 %
etre
politique
seul
president
chose
question
accord
rappeler
mauvais
besoin
vrai
fois
entendre
sentir
gagner
massacre
affaire
petain
ete
reponse
mot
hollande
apprendre
classe 5
10 %
ca
aller
tt
meme
nabilla
voir
ferie
vraiment
penser
passer
parler
con
putain
monde
hashtag
foutre
merde
pauvre
gens
gueule
aimer
gros
tellement
tweet
regarder
classe 10
6.7 %
concert
black
annuler
rappeur
maire
droite
chanter
annulation
raciste
homophobe
extreme
rap
inviter
choisir
polemique
gauche
blackm
insulter
choix
antisemite
chanteur
haine
kouffars
ps
traiter
classe 11
9.7 %
hommag
oublier
pensee
battre
liberte
rendre
patrie
heros
france
paix
sacrifice
souvenir
honneur
vie
combattre
mort
combattant
tirailleur
sacrifier
pays
homme
soldat
ancien
respect
classe 13
8.8 %
allemand
obus
batterie
gerbe
attaque
bombardemen
artillerie
troupe
nuit
offensive
depot
tir
village
tranchees
coquelicot
canon
tranchee
tirer
calme
londres
pluie
arc
sud
neige
classe 9
memoire
anneau
defi
indexations
indexer
participant
homme
objectif
tourisme
realisees
internaute
module
ficher
transcription
annotation
collaboratif
tester
record
franchir
abonnes
enrichir
saisie
indexation
see
transmission
classe 4
11.4 %
photo
gt
site
beau
decouvrez
projet
tres
nouveau
article
lire
travail
decouvrir
film
video
histoire
dernier
eleves
excellent
image
lecture
classe 3
6 %
expo
musee
exposition
visite
ouvrir
conference
soir
memorial
week_end
ee
rencontre
centre
conseil
18h
virtuel
ouverture
ees
public
visiter
rue
colloque
decouverte
art
classe 8
9.1 %
grand
guerre
ligne
front
arriere
pere
matricule
veille
cc
militaire
registre
recherche
femme
archive
retrouver
page
selection
source
breton
enealogie
trace
orient
collecte
portrait
nom
classe 12
10.6 %
bataille
ceremonie
centenaire
armistice
novembre
matin
direct
ceremonies
debut
fin
emoire
champ
emoration
mondial
speciale
anniversaire
monument
premiere
defile
suivre
commemoration
place
emission
commemorative
classe 14
indexes
livre
commun
poilu
erault
partir
carte
ere
deja
maj
nes
heraultais
integralement
ome
mam
indexees
alfort
lo
oneetloire
indexee
annotees
etaient
indexe
indexation
collaborative


MakingHistoryII -
guardian -
RachelDNA -
rightnowio_feed -
LivesOfWW1 -
I_W_M -
HistoryNeedsYou -
BBCNews -
WWIAfrica -
bbcww1 -
IWM_Centenary -
Telegraph
sommecourt -
thehistoryguy -
martin -
letter1418 -
PoppyLegion -
DCMS -
Mission1418 -
1418NOW -
virginsoldiers -
1J1Poilu -

GENÈSE DU PROJET

UN NON-PROJET DE RECHERCHE

- Mise au point d'un savoir-faire hors de mes intérêts de recherche
  - Collecter les tweets des colloques (para-recherche) (2012)
  - Echouer à collecter des tweets massivement (hors recherche) (2012)
  - Apprendre à collecter des tweets massivement (hors recherche) (2013)
- Sérendipité
  - Le rôle de socialisation des déplacements en train pour la communauté des chercheurs (RVH Blois, 2013)
  - Une recherche qui commence hors d'une intention de recherche (Avril 2014)

UN NON-PLAN DE GESTION DE DONNÉES

- Aucun plan de gestion de données

- Collecte de données artisanale (mais il y a pire)
  - Serveur configuré à domicile
  - Interruptions dans la collecte (coupure d'électricité)
  - Sauvegarde des données à la main

- Professionnalisation à l'arrivée à Lausanne
  - Passage sur un serveur de l'Unil
  - Petite redondance des données, mais qui a été résolue

UNE NON-LINÉARITÉ DE CARRIÈRE

- En dix ans
  - Unistra / CVCE (Luxembourg) / Paris-Sorbonne / Unil
  - Depuis 2014:
    Paris-Sorbonne (gestion *lolesque* des données) / Université de Lausanne (poste temporaire)
  - Ne favorise pas la mise en place d'une gestion de ses données liées à la recherche

UN PLAN DE GESTION DES DONNÉES AURAIT-IL FAVORISÉ CETTE RECHERCHE?

- Non
  - Tâche de recherche, certes, mais administrative aussi
  - Ne facilite pas la sérendipité, l'empêche peut-être
- Adapté aux grands projets de recherche, nécessitant financement national / européen / international
  - Pas aux petits projets de recherche
  - Pas aux projets de recherche liés à la sérendipité
  - Ces petits projets de recherche font les grands projets

Le plan de gestion de données ne doit pas être une machine à empêcher la créativité de la recherche et la sérendipité, c'est-à-dire ce que la recherche doit au hasard

ET AUJOURD'HUI?

- Oui
  - Il faut adapter la notion de plan de gestion de données à la diversité des modes de recherche
  - Aujourd'hui, ce projet de recherche nécessite un plan de gestion des données, notamment en vue de leur conservation
  - Difficulté liée aux changements de statut / d'université
- Pour l'Unil et les SHS, la plateforme technique du LaDHUL (PlaTec - https://platec.unil.ch/) répond à ce besoin de diversité