

Exploitation d'une base de données

SAE S2.04



Table des matières

Introduction	2
Page d'accueil	3
Rappel de la demande :	3
Choix de développement :	3
Le code :	4
Le résultat :	4
Page Corpus	5
Rappel de la demande :	5
Choix de développement :	5
Le code :	6
Le résultat :	6
Page Exploration	8
Rappel de la demande :	8
Choix de développement :	8
Le code :	<u>c</u>
Le résultat :	<u>9</u>
Page Temporalité	10
Rappel de la demande :	10
Choix de développement :	10
Le code :	10
Le résultat :	10
Page Retweets	11
Rappel de la demande :	11
Choix de développement :	11
Le code :	11
Le résultat :	11
Page Activité	12
Rappel de la demande :	12
Choix de développement :	12
Le code :	12
Le résultat :	12
Conclusion	13
Bibliographie	14

Introduction

Durant cet SAE nous devions réaliser une application à partir d'un cahier des charges et d'une base de données fournie par le client. Pour cela nous devions répondre à certaines exigences imposées par celui-ci et donc réaliser une application propre à l'attente du client. Le client réalise des études d'opinion à partir de données Twitter moissonnées avec DMI-TCAT. L'étude qui nous a été confiée concerne les tweets de personnalités politiques sur la période de 2012 à 2018. Quant à l'application web, elle a été réalisée sur la plateforme Oracle Apex.

L'application web s'organise sur 6 pages :

- Une page d'accueil
- Une page Corpus
- Une page Exploration
- Une page Retweets
- Une page Activité

Chacun de ces 6 pages possède des caractéristiques et rôles qui leur sont propres avec pour chacune d'entre d'elles, certaines exigences et attentes de la part du client.

Quant au rapport, vous trouverez la description de chacune des pages présentes dans l'application avec pour chacune d'entre elles, une description détaillée avec un rappel de la demande concernant la page en question, mais aussi du choix concernant les composants choisis et de la raison de ces choix, avec la description du code éventuellement effectué et pour finir, une capture d'écran montrant le résultat, soit le visuel des différentes pages.

URL de l'application :

https://apex.oracle.com/pls/apex/r/r_206-2022-hmn4425a/twitter-dat1/login

Page d'accueil

Rappel de la demande :

Concernant la 1ère page, il nous a était demandé de réaliser une page d'accueil comportant 3 parties :

- Un titre où doit figurer mon nom et prénom, dans la phrase « Dashboard développé par nom prénom. »
- Une phrase explicative concernant le tableau de bord, composée de variables qui sont à calculer à partir des données et qui reflètent ainsi tout ajout ou suppression de tweets dans la base de données.
- Nous devions par la suite réaliser un diagramme à barres. Ce diagramme montre le nombre de tweets postés par personne, les plus volubiles étant à gauche de l'axe des abscisses.

Choix de développement :

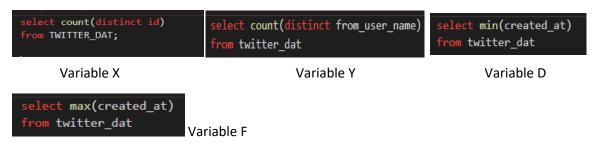
Pour de ce qui est des choix de composants choisis pour la réalisation de cette page, nous nous referions aux demandes du client. Pour la phrase explicative, mettre du contenu statique dans la partie fil d'Ariane répond au mieux aux exigences du client, ce contenu statique prend en code HTML, la phrase voulut composer de variables qui varieront selon la base de données et à laquelle sera assigné un code SQL qui permettra de leur donner une valeur précise de cette base. Nous avons ainsi créé 4 variables X, Y, D et F, qui auront toutes une valeur chiffrée.

Par la suite nous devions aussi créer un diagramme à barres, dont les exigences étaient données, il n'y avait là pas réellement de choix judicieux à mettre en œuvre mais seulement respecter les attentes du client. Nous avons ainsi inséré dans le body, soit le corps de la page un élément graphique, qui par défaut représente un diagramme à barres, entre autres l'exigence du client.

Le code:

Quant au code effectué, il s'agit de requêtes assignées aux différentes variables, ce sont des requêtes SQL sur la base de données implémentée dans l'application, les requêtes ont en soit la même structure mais ont des objectifs différents. En effet la variable X compte le nombre de tweets émis par les personnalités politiques qui eux sont comptés dans la variable Y et pour les variables D et F, elles sont des dates qui correspondent à une certaine période, soit là, pour la variable D la date minimum et pour la variable F la date maximum. À la suite de la création de ces différentes variables, il nous fallait les implémenter dans une phrase déjà définie.

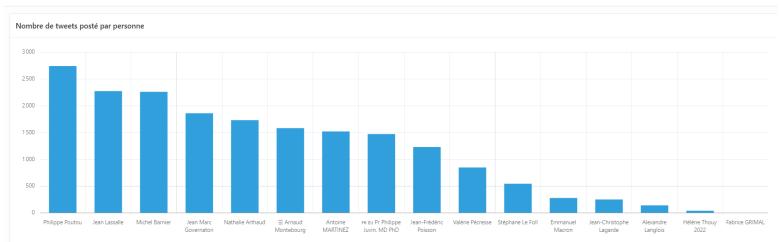
Pour le diagramme à barres, nous avions à mettre dans les sources du diagramme, la base de données fournie par le client et de sélectionner les différents éléments demandés, dans notre cas le nom complet de toutes les personnalités politiques et du nombre de tweets qui leurs sont propres, tout ça en ordre décroissant comme exigé par le client.



Le résultat :

Dashboard développé par Hmadouch Anass

Ce tableau de board permet d'explorer les 18816 tweets de 16 personnalités politiques, moissonnés entre le 04/06/2012 et le 31/12/2018.



Page Corpus

Rappel de la demande :

Pour la réalisation de la page corpus, il nous est demandé de présenter un rapport dynamique, initialement trié par Date, incluant 8 colonnes :

- Date avec comme format la date complète (DD/MM/YYYY), l'heure et les minutes
- Personne avec un lien pointant sur le profil Twitter de la personne
- Certifié contenant 0 (non certifié) ou 1 (certifié)
- Tweet avec le contenu du message posté
- Retweets avec le nombre de retweets que ce tweet a suscité
- Favoris avec le nombre de favoris que ce tweet a suscité
- Dispositif avec le nom du dispositif utilisé pour publier le tweet
- Lieu avec la localisation fournie par Twitter

Pour donner suite à la réalisation de 7 états publics, représentant en quelque sorte des filtres selon certains critères définis et choisis par le client.

Ces différents états sont :

- État principal : le rapport trié par Date
- Dispositifs : la liste des Dispositifs et le nombre de personnes ayant utilisé chacun, trié sur la 2e colonne.
- Thème Occitanie: la recherche du mot « Occitanie » dans la colonne Tweet.
- Thème solidarité : la recherche du mot « solidarité » dans la colonne Tweet.
- Top favoris : le rapport trié par Favoris décroissants, puis par Retweets décroissants en cas d'ex aequo.
- Top retweets : le rapport trié par Retweets décroissants, puis par Favoris décroissants en cas d'ex aequo.
- Toulouse et Montpellier : les Tweets mentionnant Toulouse et les Tweets mentionnant Montpellier.

Choix de développement :

Pour de ce qui est du choix des composants pour la page corpus, le choix de mettre un aspect rapport dynamique, représente le choix le plus judicieux à mettre en place. Ce rapport est en quelque sorte un tableau interactif répertoriant tous les éléments de la base de données. Cependant, nous est demandé d'y mettre seulement certains éléments, pour cela nous faisons un tri, sélectionnant ainsi les éléments selon les exigences du client.

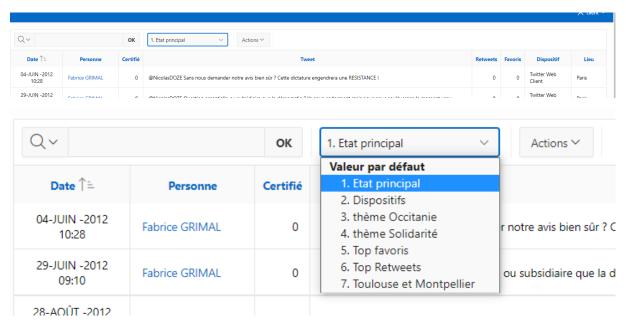
Par la suite, était à mettre en place les différents états listés plus haut, pour cela nous avions à créer des filtres pour chaque état et à les enregistrer en tant qu'états sur la page corpus elle-même sans passer par les fonctionnalités d'Apex Oracle.

Le code:

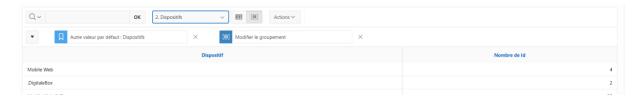
Pour la page corpus, certains éléments étaient demandés et d'autres non, de ce fait, étaient à modifier certains éléments pour obtenir le résultat attendu, les libellés étant aussi non conformes à l'attente du client, ils ont donc été modifiés. En soit pour la sélection des 8 colonnes, 2 choix pouvait s'offrir à nous, l'un d'eux est le tri de colonne dans le corpus, en sélectionnant celles voulues, puis la possibilité de supprimer ou encore masquer les colonnes non demandées dans l'App Builder. De plus, il nous a était demandé de créer un lien pour chaque personnalité politique dès lors que nous cliquons sur leurs noms dans la colonne « Personne », pour cela nous avons simplement mis le type de la colonne en tant que lien et avons fait une sorte de concaténation avec l'URL twitter et l'identifiant des ces personnalités politiques, qu'on retrouve dans la colonne FROM USER NAME.

Le résultat :

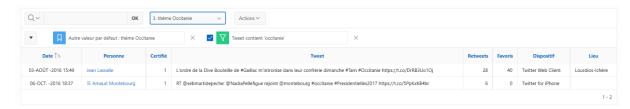
Etat principal:



Dispositifs:



Thème Occitanie:



Thème Solidarité:



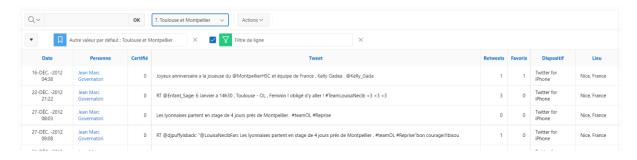
Top favoris:



Top Retweets:



Toulouse et Montpellier :



Page Exploration

Rappel de la demande :

Concernant la page Exploration, il nous était demandé de présenter un rapport à facettes, avec comme pour la page corpus certain éléments. On retrouve alors les colonnes date, personne, certifié, tweets, retweets, favoris, dispositif et lieu. Quant aux facettes à réaliser, elles sont en nombre de 7 dont l'ordre est à respecter :

- Personne, qui répertorie toute les personnalités politiques
- Certifié, qui affiche 1 quant la personnalité politique est certifiée par Twitter ou sinon affiche 0
- Langue, qui correspond aux différentes langues des tweets
- Dispositif, qui répertorie le nom des différents dispositifs utilisé pour publier le tweet
- Localisation avec les différentes localisations fournies par Twitter
- Retweets, qui répertorie les différentes possibilités concernant le nombre de retweets qu'un tweet a pu susciter.
- Favoris avec le nombre de favoris qu'un tweet a pu faire.

Choix de développement :

Tout comme pour la page corpus, nous avions à faire un rapport mais à la différence de celui-ci, il était là question d'un rapport par recherche à facettes qui donne aux utilisateurs les moyens de filtrer une collection de données en choisissant une ou plusieurs facettes. Elle consiste à lister en précisant le nombre de résultats répondant à la fois à la requête effectuée et à la facette. La sélection d'une facette supplémentaire va affiner la requête courante en appliquant le critère lié à la facette en plus des critères déjà sélectionnés, pour ne présenter que les résultats issus de la réunion de toutes ces facettes.

Ces différentes facettes agiront sur un rapport, dont le choix des colonnes est le même que pour la page Corpus.

Le code:

Pour la page Exploration, étaient demandées les mêmes colonnes que pour la page corpus, nous avons donc fait les mêmes manipulations pour arriver au même résultat. En outre ici, masquer celles qui ne sont pas cités plus haut et démasquer celles voulus. Puis nous rajoutons à la page un aspect recherche par facette que nous plaçons à gauche du rapport et dont nous ajoutons autant de facettes que ceux exigées par le client. Les informations présentes dans ces facettes sont, tout comme la plupart des informations des différentes pages, issues de la base de données fournie par le client. Nous devons ainsi faire le lien entre la facette et l'élément en question de la base de données.

Q Rechercher		Nombre total de lignes 19 k						
∨ Personne	0	Date ↑=	Personne	Tweet	Retweets	Favoris	Dispositif	Lieu
Philippe Poutou (2 744) Jean Lassalle (2 276) Michel Barnier (2 263) Jean Marc Governatori (1 863) Nathalie Arthaud (1 734) Afficher tout		04-JUIN -2012 10:28	Fabrice GRIMAL	@NicolasDOZE Sans nous demander notre avis bien sûr? Cette dictature engendrera une RESISTANCE !	0	0	Paris	Twitter Web Client
		29-JUIN -2012 09:10	Fabrice GRIMAL	@NicolasDOZE Question essentielle ou subsidiaire que la démocratie ? Ils nous endorment mais nous nous soulèverons le moment venu.	0	0	Paris	Twitter Web Client
		28-AOÛT -2012 16:15	Fabrice GRIMAL	@NicolasDOZE "Plus d'Europe ou le chaos" répété en boucle dans votre émission des experts d'hier. CHANTAGE ! Intolérable en démocratie.	0	0	Paris	Twitter Web Client
	۰	15-OCT2012 17:21	Jean Marc Governatori	L' @OLfeminin c'est 5 matchs. 5 victoires. 40 buts marqués. 0 encaissé !	4	3	Nice, France	Twitter for iPhone
> Certifié	8	16-OCT2012 09:59	Jean Marc Governatori	ITW de Sonia Bompastor http://t.co/4AP5tzFf	0	0	Nice, France	Twitter for iPhone
> Langue	0	16-OCT2012 15:21	Jean Marc Governatori	RT @fifi5035: @LouisaNecibFan demain je vais à Clairefontaine assister à l'entraînement	1	0	Nice, France	Twitter for iPhone
Dispositif	0	16-OCT2012 20:45	Jean Marc Governatori	RT @Akhi5830: L'equipe de france feminine est quand meme plus impressionnante @LouisaNecibFan @laureboulleau	1	0	Nice, France	Twitter for iPhone
> Retweets	0	17-OCT2012 20:13	Jean Marc Governatori	RT @UneDeuxNet: #Bompastor en interview, très enrichissant. A lire demain sur http://t.co/UvAaVHfl, faites passez l'info! #EDFféminine	16	0	Nice, France	Twitter for iPhone
> Favoris	0	18-OCT2012 20:42	Jean Marc Governatori	@LouisaNecibFan et le compte Facebook LouisaNecib #14 s'associent pour vous offrir de meilleures infos photos vidéo, sur Louisa Necib.	0	0	Nice, France	Twitter for iPhone
		19-OCT2012 22:05	Jean Marc Governatori	Qui sera au match demain soir a Charlety ? #EDFféminine #FranceAngleterre	1	0	Nice, France	Twitter for iPhone
		19-OCT2012 22:12	Jean Marc Governatori	@Allicee80 @_maryie De vrais supporters J'y serais aussi faudra encourager Louisa Je compte sur vous ;)	0	0	Nice, France	Twitter for iPhone
		19-OCT2012 22:16	Jean Marc Governatori	Demain je serais a Charlety! Les supporters de Louisa j'espère vous entendre !! Mettez l'ambiance ! (Toto on essavera de ce canter ! ;))	1	0	Nice, France	Twitter for iPhone

Page Temporalité

Rappel de la demande :

Pour la page Temporalité, il nous est demandé de présenter un diagramme à barres par mois croissants, avec seulement le mois et l'année, représentant le nombre de tweets publiés chaque mois.

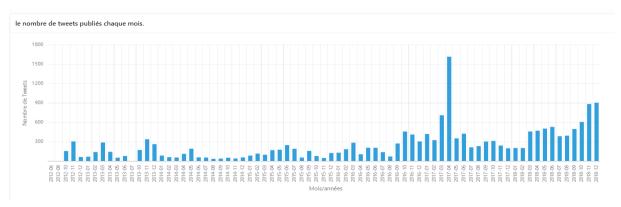
Choix de développement :

Concernant le choix des composants, nous avons seulement inséré dans le corps de la page un graphique de type diagramme à barres et nous y avons importé la base de données pour pouvoir ainsi répondre à la demande du client, soit le nombre de tweets publié par mois. Le choix du diagramme à barres, dépend notamment de son aptitude à montrer l'évolution d'une variable sur une période définie, ce qui dans notre cas représente notre objectif.

Le code:

Quant aux réglages de propriété du composant effectué, il nous fallait avoir le nombre de tweets publié par mois, pour cela nous devions réaliser une requête SQL et donc modifier la source du composant en passant de table/vue pour le type, à Requête SQL et pouvoir ainsi rédiger notre requête. Quant à la requête en elle-même, il nous fallait afficher seulement le mois et l'année et aussi le nombre de tweets du mois et année concernées que nous groupons par date et affichons en ordre croissant. Suite à cela, est affiché le libellé des deux axes, en donnant un titre à l'axe X et Y.

```
select to_char(created_at, 'YYYY-MM') as mm_yyyy, count(text) as nbTweet
from twitter_dat
group by to_char(created_at, 'YYYY-MM')
order by to_char(created_at, 'YYYY-MM') asc;
```



Page Retweets

Rappel de la demande :

Il nous est demandé pour la page Retweets, de présenter un diagramme montrant le nombre de retweets obtenus par les tweets postés par mois (seulement le mois et l'année) pour 3 personnalités politique : Nathalie Arthaud, Jean Lassalle et Philippe Poutou, avec pour chacun, une couleur pour les distinguer.

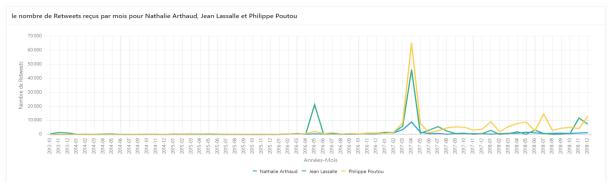
Choix de développement :

Pour le choix du composant, il est nous est demandé de faire un graphique en courbe permettant de représenter les données sur la période de 2012 à 2018. L'axe horizontal présente une progression continue dans le temps. L'axe vertical quant à lui présente l'évolution des valeurs au fil de cette progression. Entre autres, ce graphique est très pratique pour visualiser l'évolution des valeurs dans le temps. De plus, il nous faut rajouter une légende pour identifier les différents acteurs du graphique.

Le code :

Quant aux réglages de propriété du composant effectué, il nous fallait dans un 1^{er} temps créer un diagramme à courbes qu'on retrouve dans les attributs du graphique. A la suite de cela, nous créons 3 séries pour les 3 personnalités politiques (Nathalie Arthaud, Jean Lassalle et Philippe Poutou) et pour chacun d'eux, on y met une requête SQL pour chacune des 3 séries qui sélectionne la date (le mois et l'année seulement), depuis la base de données où nous comptons le nombre de tweets distincts pour chacune des 3 personnalités. Que nous trions dans un ordre croissant.

```
Select to_char(created_at, 'YYYY-MM'), sum(retweet_count)
from TWITTER_DAT
where FROM_USER_NAME = 'n_arthaud'
group by to_char(created_at, 'YYYY-MM')
order by to_char(created_at, 'YYYY-MM') desc
```



Page Activité

Rappel de la demande :

Pour finir, nous devions réaliser une page activité, présentant une boîte à moustaches par an résumant la distribution du nombre de retweets que les publications de Philippe Poutou ont reçu.

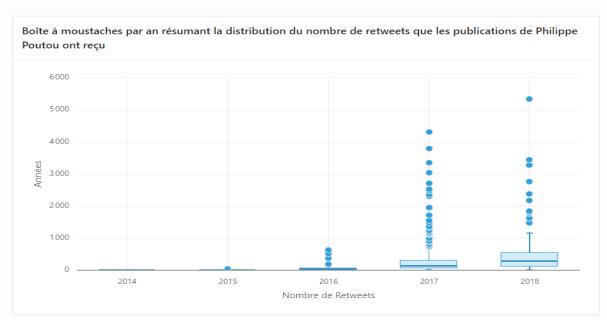
Choix de développement :

Quant au choix du composant, il s'agit là de réaliser une boîte à moustache qui va nous permettre d'identifier les points de données d'intérêt ou les valeurs dites « aberrantes », c'est-à-dire des valeurs qui contrastent grandement avec les valeurs normalement mesurées. Ces valeurs sont tracées comme des points de données et tombent au-delà des moustaches.

Le code:

Concernant les réglages de propriété effectués, tout comme pour les autres graphiques, il nous faut paramétrer le graphique pour qu'il puisse correspondre à une boîte à moustaches. Suite à cela, il ne nous reste plus qu'à y écrire une requête SQL à partir de la base de données fournie par le client et d'y grouper par année le nombre de retweets que les publications de Philippe Poutou ont reçu et seulement lui.

```
Select to_char(created_at, 'YYYY'), retweet_count as nbTweets
from TWITTER_DAT
where FROM_USER_NAME = 'PhilippePoutou'
group by to_char(created_at, 'YYYY'), retweet_count
order by to_char(created_at, 'YYYY') desc
```



Conclusion

Durant cette SAE, bon nombre de compétences nous ont été requises pour la réalisation de celui-ci. En effet mise à part l'aspect quelque peu intuitif de Apex Oracle cela nous a demandé un minimum de réflexion et de recherches documentaires à son sujet, pour mieux comprendre les fonctionnalités et ainsi répondre au mieux aux besoins du client.

Mise à part cela, quelques autres connaissances nous ont été requises, comme une base solide quant à l'aspect SQL du travail, car il ne s'agit pas seulement de rédiger des requêtes mais il s'agit là de répondre à une demande et donc d'identifier les différents éléments à mettre en place pour avoir le meilleur rendu possible, mais il s'agit surtout d'y rédiger des requêtes sur un environnement différent de ceux présentés jusqu'à présent et d'avoir aussi un certain niveau en anglais. Il nous est de plus demandé d'avoir des connaissances quant à l'aspect graphique, des différents agencements des composants de la page, avec pour nombre d'entre eux un paramétrage propre à chacun. En plus d'une certaine connaissance sur les différents diagrammes et de leurs utilités ainsi que des situations où ils sont le plus propice à mettre en œuvre.

Bibliographie

Durant la réalisation de cette application et de ce rapport, bon nombre de ressources ont été consulté pour parvenir aux mieux au résultat attendu. Pour la réalisation de l'application, elle a été élaborer sur l'Environnement de développement Apex Oracle, environnement qui reste quelque peu compliqué à prendre en main dans un premier temps, du fait de ses nombreuses fonctionnalités.

Mais forte heureusement pour nous, nous avions à notre disposition un cours sur l'utilisation d'Apex mais aussi d'un TP que nous avons réaliser au préalable et qui nous à ainsi permis de retrouver quelques points majeurs que l'en retrouve dans le sujet de SAE.

Mise à part cette documentation-ci, quelque question se poser sur les différents composants à mettre mais surtout de leurs utilités. En effet, pour répondre aux demandes et exigences d'un client, il faut savoir quels composants peut répondre au mieux à ces demandes. Pour cela était à notre disposition des liens nous ramenons vers un bon nombre de documentation explicative des composants et fonctionnalité d'Apex.

De plus, des recherches quant à certain éléments ont réaliser, pour comprendre aux mieux leurs utilités, tels que pour la réalisation des graphiques et ainsi comprendre leurs utilités et les différents contextes où ils seront les plus propice à mettre en place. Pour cela, de simple documentation à leurs sujets suffise, on retrouve Wikipédia, mais d'autre sites spécialiser en statistique, tel que www.imp.com.