



AL-KHWARIZMI

INSTITUTE OF SCIENCE, TECHNOLOGY AND
INNOVATION

RAPPORT DE PROJET

Développement d'une application web de la recommandation de news

Realisé par:

Brahim ERRAJI

Encadré par:

Pr. Rachad TAOUFIK

Année Universitaire 2019/2020

Remerciements

Je tiens à remercier le Professeur **Rachad Taoufik** pour son soutien tout au long la periode de réalisation de ce projet.

Contents

1	Introduction	3
2	Présentation du projet	4
2.1	Problématique	4
2.2	Solution proposée	4
2.3	Objectifs du projet	4
3	Analyse et Conception	5
3.1	Collecte de données	5
3.1.1	Le scraping	5
3.1.2	API	6
3.2	Base de données	7
3.2.1	Les tables de la données	7
3.2.2	Les associations	7
3.2.3	Schéma	8
3.3	La recommandation	9
3.4	L’affichage des articles	10
4	Réalisation	10
4.1	Architecture de l’application web	10
4.2	Arborescence de l’application web	11
5	Démonstration	12
5.1	Page d’inscription	12
5.2	Page de connexion	13
5.3	Page d’accueil	13
5.4	Page de catégorie	14
5.5	Page d’article	14
5.6	Page d’articles sauvegardés	15
6	Outils de réalisation	16
7	Conclusions et perspectives	17
	□	

1 Introduction

Dans le cadre de la mise en pratique des connaissances acquises durant les deux premières années de la licence science des données, 16 projets individuels ont été affectés aux étudiants de la promotion 2018/2019.

Le projet qui m'a été affecté est le *Développement d'une application web de la recommandation de news*, un projet proposé par Professeur Rachad Taoufik qui touche à tout aspect de la formation (Conception des applications web, traitement des données, gestion de projet ..).

2 Présentation du projet

2.1 Problématique

Dans les sites Web d'informations traditionnels, l'utilisateur doit :

- Lire tous les articles disponibles du jour
- Décider manuellement ceux qui l'intéressent

ce sont des tâches fastidieuses et longues et l'utilisateur risque de ne pas voir des articles qui pourraient être très intéressants pour lui. créant ainsi le besoin d'un site de news qui parcourt les articles venant de différentes sources et ne montre que ceux qui sont pertinents à l'utilisateur.

2.2 Solution proposée

Pour répondre à ce besoin, on a besoin d'une application web qui fait ces tâches à la place de l'utilisateur, c-a-d elle consulte régulièrement les sites de news connus et en extrait les articles puis elle les stocke dans une base données qui sera exploitée pour recommander des articles à chaque utilisateur. La base de donnée contiendra également des informations sur l'utilisateur (personnel et comportemental).

2.3 Objectifs du projet

L'application web doit être capable de :

- Parcourir les sources de news connues et récupérer tous ses articles
- Stocker les articles dans une base de données
- Exécuter l'algorithme de recommandation
- Afficher les articles pertinents à chaque utilisateur

3 Analyse et Conception

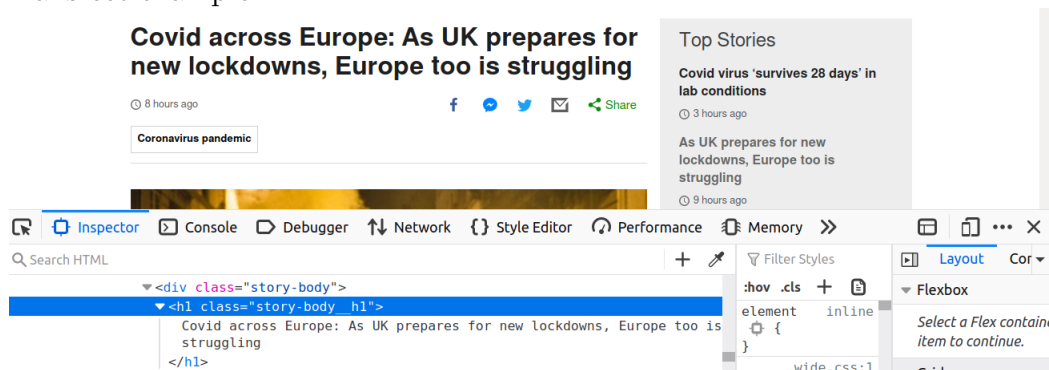
3.1 Collecte de données

Pour collecter les articles, il y a deux méthodes :

3.1.1 Le scraping

Le *web scraping* est une technique d'extraction du contenu de sites Web, via un script ou un programme, dans le but de le réutiliser différemment. Il se base sur les balises HTML pour trouver le contenu désiré.

Dans cet exemple:



pour extraire le titre de l'article on peut créer un script qui cherche dans le code source de la page une balise HTML de type `"h1"` qui a `"story-body h1"` comme class et en extrait le texte, on tombe alors sur le titre de l'article et ainsi de suite pour les autres informations.

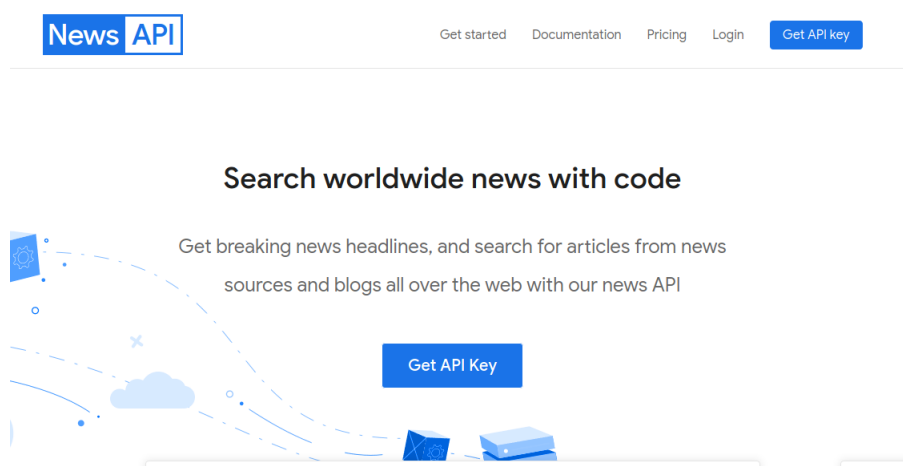
cette méthode permet d'extraire les informations d'un article et les rendre facile à manipuler pour les soumettre dans une base de données.

Même si le scraping donne une liberté de manipulation, il présente néanmoins des difficultés :

- Le script de scraping ne marche pas pour tous les sites, même pas pour toutes les pages au sein du même site, ce qui nécessite un script pour chaque page.
- Le code source d'une page peut être changé à tout moment, et dans ce cas le script initial ne marche plus et il faut en créer un autre.

3.1.2 API

Une méthode alternative du scraping est l'utilisation d'un **API** de news. Un **API** est l'équivalent d'un serveur en informatique, après la réception d'une requête il renvoie une réponse contenant les informations demandées. Un API de news renvoie les articles sous forme de réponse JSON. Dans ce projet L'API **NEWS API** sera utilisé.



3.2 Base de données

3.2.1 Les tables de la données

Les tables principales de la bdd sont :

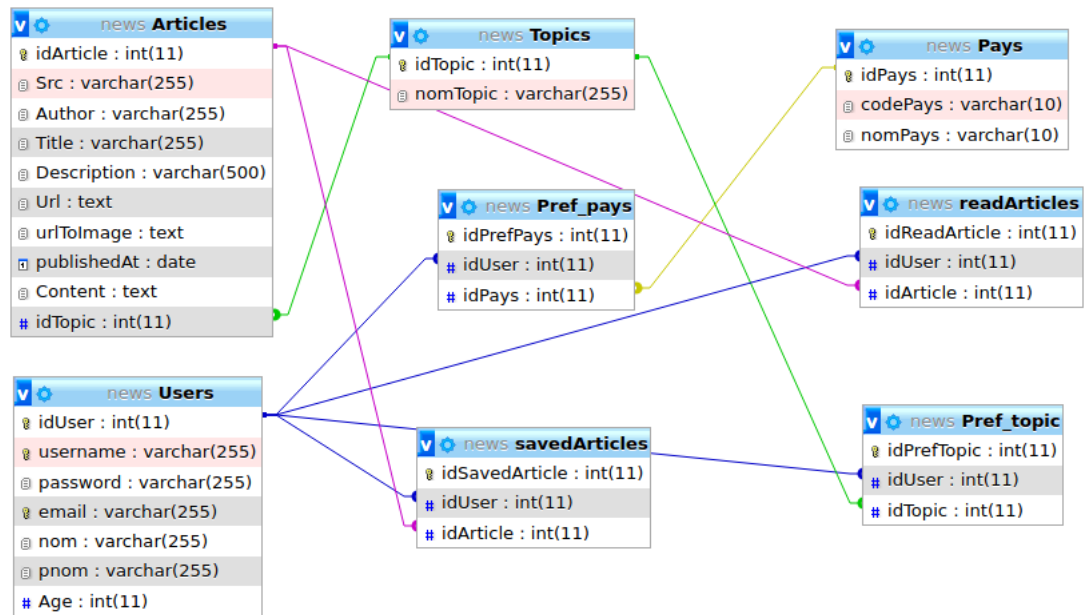
- Articles
- Users
- Topics
- Pays

3.2.2 Les associations

Dans La base de données, les associations suivantes sont implémentées :

- Un utilisateur peut préférer plusieurs catégories (Topics)
- Un utilisateur peut préférer plusieurs pays
- Un utilisateur peut lire plusieurs articles
- Un utilisateur peut sauvegarder plusieurs articles
- Un article appartient à une seule categorie (Topic)

3.2.3 Schéma



3.3 La recommandation

Après la collecte des articles, L'application donne un score de pertinence à chaque article par rapport à chaque utilisateur. Ce score est calculé en répondant aux questions suivantes :

- la catégorie de l'article relève-t-elle des intérêts de l'utilisateur?
- L'article ressemble-t-il à des articles précédemment lus par l'utilisateur?
- l'article est-il tendance dans le pays de l'utilisateur? (Des utilisateurs du même pays ont exprimé leur intérêt pour l'article)
- ...

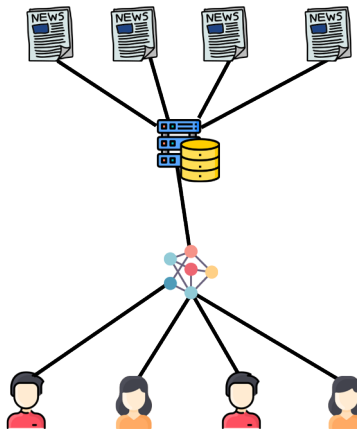
3.4 L’affichage des articles

La dernière phase consiste à afficher les articles jugés plus pertinents aux utilisateurs. Chaque page d’article contient

- Le titre
- Une image
- Une petite description
- Le lien de l’article complet

4 Réalisation

4.1 Architecture de l’application web

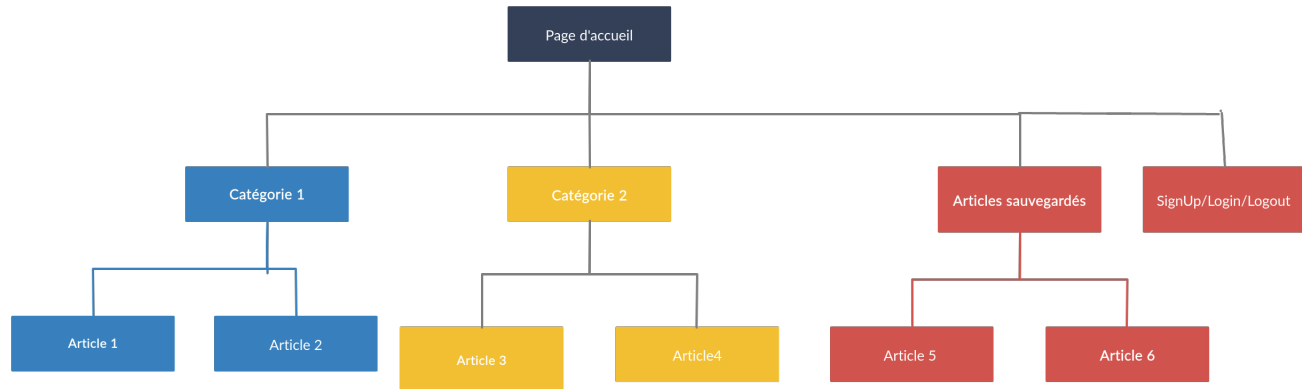


Pour faire son travail, l’application passe par trois phases principales :

- Le scraping ou bien la collecte des articles d’actualités
- Injection des articles dans la base de données
- La recommandation ou bien trouver les articles pertinents à chaque utilisateur
- L’affichage des articles aux utilisateurs respectivement

4.2 Arborescence de l'application web

Après que l'utilisateur est redigiré vers la page d'accueil.



5 Démonstration

5.1 Page d'inscription

BOXNEWS[Sign up](#)[Sign in](#)

Sign Up

Please fill in this form to create an account.

Email

Enter Email

Username

Enter Username

Password

Enter Password

First Name

Enter First Name

Last Name

Enter Last Name

Age

Enter Age

Select topics that you're interested in :

Business

Entertainment

General

Health

Science

Sports

Technology

Cancel

Sign Up

Copyright © 2020 BoxNews All Rights Reserved.

5.2 Page de connexion

BOXNEWS [Sign up](#) [Sign in](#)

Sign In

Username

Enter Username

Password

Enter Password

Cancel

Sign In

Copyright © 2020 BoxNews All Rights Reserved.

5.3 Page d'accueil


BOXNEWS [Saved articles](#) [Business](#) [Entertainment](#) [General](#) [Health](#) [Science](#) [Sports](#) [Technology](#) [User1](#) [Log out](#)

Look for an article

From:

To:


Submit



The Crown series four FIRST LOOK: Emma Corrin is the spitting image of Princess Diana - Daily Mail

Actress Emma Corrin was the spitting image of the Princess of Wales in a newly released image posted to the show's official Twitter account on Saturday.


[Read more](#)



Chrishell Stause is red hot in a strapless gown as she arrives at DWTS rehearsals - Daily Mail

Chrishell Stause, 39, arrives to rehearsal for Dancing With The Stars in a red gown as she greets her partner Gleb with a hug and a million dollar smile as he carried coffee and breakfast for the pair


[Read more](#)





'Mean Girls' Cast Reunites After 16 Years to Encourage Voting - Variety

With Election Day one month away, Tina Fey, Lindsay Lohan, Rachel McAdams, Amanda Seyfried and the cast of "Mean Girls" released a surprise reunion video encouraging fans to vote. The virtual reunion, moderated by Katie Couric and released via Instagram, also...

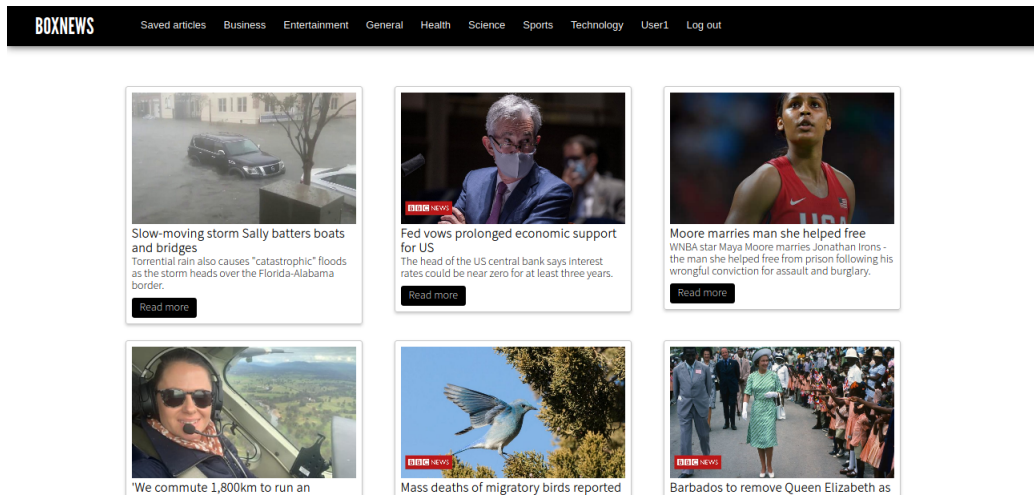
[Read more](#)



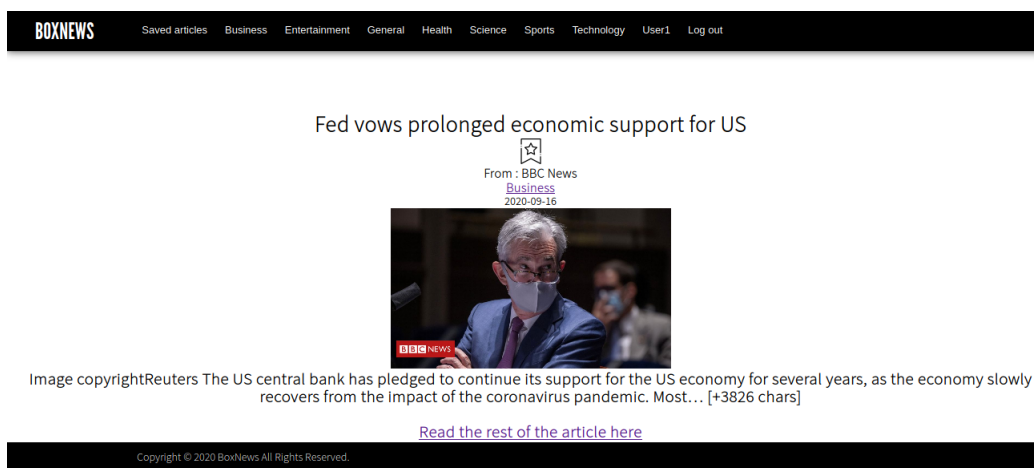




5.4 Page de catégorie




5.5 Page d'article



5.6 Page d'articles sauvegardés

BOXNEWS [Saved articles](#) [Business](#) [Entertainment](#) [General](#) [Health](#) [Science](#) [Sports](#) [Technology](#) [User1](#) [Log out](#)

Fed vows prolonged economic support for US


From : BBC News
[Business](#)
2020-09-16




Image copyrightReuters The US central bank has pledged to continue its support for the US economy for several years, as the economy slowly recovers from the impact of the coronavirus pandemic. Most... [+3826 chars]

[Read the rest of the article here](#)

Copyright © 2020 BoxNews All Rights Reserved.

6 Outils de réalisation

Pour réaliser ce projet, plusieurs outils de developpement web et gestion de projet ont été exploités :

- HTML/CSS/JavaScript pour la partie Front End
- PHP/MySQL Pour La partie Back End
- IceScrum pour garder trace des diffrentes fonctionalités de l'application et des dates clés

7 Conclusions et perspectives

L'application web BoxNews est loin d'être complete, Néanmoins elle pourrait répondre à plusieurs besoins de l'utilisateur :

- Créer un compte
- Avoir une recommandation à base de la catégorie (Topic) de l'article
- Lire un article
- Sauvegarder un article
- Chercher un article
- ...

Mais il reste plusieurs aspects à améliorer :

- Ajouter la possibilité de partager l'article avec ses amis
- Ajouter la possibilité de réagir avec un article (like, dislike)
- Ajouter un espace commentaire
- Améliorer la recommandation (mieux comprendre l'utilisateur)
- ...