

<https://profdoc.iddocs.fr/spip.php?article56>

Comprendre le moteur de recherche avec une simulation

- Contenus pédagogiques -

Date de mise en ligne : lundi 15 février 2016

profdoc.iddocs.fr - prof' doc' - Creative Commons CC BY-NC-SA

Au-delà de manipulations de moteurs de recherche comme Google, Yahoo !, Bing, Qwant ou Duckduckgo, il est difficile d'aborder le lien entre utilisation d'un moteur de recherche et sa base de données. La conception d'une simulation est une tentative pour apporter cette connaissance aux élèves, dès la Cinquième.

L'utilisation pédagogique de l'outil présenté dans cet article est intégrée dans une séquence de niveau Cinquième, intitulée [Du document au média](#). Dans cette séquence, les élèves sont amenés à découvrir la chaîne éditoriale du livre, les étapes de conception d'une page web, puis le fonctionnement d'un moteur de recherche pour trouver les pages web.

Initialement développée pour le niveau de Cinquième, cette simulation a évolué pour être utilisée selon trois niveaux, ainsi en fin de cycle 3 ou début de cycle 4 (10-12 ans), en fin de cycle 4 (13-15 ans) et au lycée (16-18 ans). De même, d'abord prévue pour une séance d'une heure, elle est maintenant déclinable pour 2 heures ou pour un travail en autonomie.

La version 2.0 a été développée à partir des remarques rédigées par Julien Rousseau dans son article *Webfinder : le nez dans le moteur*, publié sur le site *Les Trois Couronnes* [1]. Merci par ailleurs à Olivier LeDeuff d'avoir suggéré l'idée d'une version anglophone, intégrée donc dans la version 2, avec par ailleurs la possibilité d'autres langues par la suite.

La simulation se présente sous la forme d'un site web, **Webfinder**, accessible sur : <http://iddocs.fr/webfinder/> Ce site web est disponible ci-dessous en téléchargement, afin de l'améliorer et de l'héberger ailleurs.

Précisons enfin que la simulation existe en deux langues, français et anglais, avec des pages web référencées selon chaque langue.

Utilisation pédagogique

La simulation *Webfinder*, pour favoriser la compréhension du fonctionnement du moteur de recherche, part du principe qu'on ne compte que dix pages web sur Internet. Le moteur de recherche est créé pour permettre la recherche parmi l'ensemble de ces pages, en estimant qu'il y en aura davantage ensuite (plus de mille milliards aujourd'hui).

Le travail passe d'abord par la consultation de la base de données, avec un tableau de dix lignes (dix pages web) et de cinq colonnes (les élèves ayant déjà été initié au langage HTML, observent et expliquent). Précisons que les nombre donnés ne sont pas réels, hormis le calcul du nombre de mots du lemme.

Passant sur la page suivante, les élèves sont amenés à faire une recherche autour de la digestion (les pages traitent toutes de la digestion), et à repérer le critère qui explique le classement des dix pages dans cet ordre précis. Il est possible de trier le tableau à partir d'une colonne en cliquant sur l'en-tête de la colonne choisie.

Cette étape est la plus longue, peut demander 15 à 20 minutes de réflexion afin que les élèves s'approprient l'outil,

l'effort de comparaison entre résultats et tableau de la base de données. Deux étapes suivantes consistent à ajouter deux critères de classement. La simulation comprend aussi une partie relative aux algorithmes, une autre au droit à l'oubli ou déréférencement, une autre à l'indexation, une dernière à la publicité.

Ces éléments sont évidemment complexes pour des élèves de Cinquième, ils sont envisagés selon plusieurs niveaux dans la version 2, et peuvent être accompagnés d'une comparaison de moteurs de recherche et de la découverte d'autres bases de données ([celles de médias sociaux en ligne par exemple](#)).

Niveaux et modes

Dans la version 2, l'utilisateur peut choisir entre trois niveaux et entre trois modes.

Pour le **niveau** débutant, la recherche des réponses est plus simple, avec des indications données qui disparaissent pour les niveaux intermédiaire et avancé.

A partir du niveau intermédiaire, il y a une partie supplémentaire sur le déréférencement. Au niveau avancé, il y a une partie supplémentaire sur l'indexation (avec une entrée par le refus d'indexation).

Dans les **modes** sur 2 heures ou en autonomie, il y a davantage de questions, ainsi au sujet des algorithmes ou de l'absence d'indexation. Par ailleurs le mode sur deux heures propose une page récapitulative de rédaction de définitions, qui n'existe pas dans le mode en classe d'une heure.

Dans le mode en autonomie, les réponses sont données après trois échecs, alors qu'en classe (quel que soit le temps de la séance), il faut que les réponses soient renseignées pour passer à la suite (si besoin donc avec l'apport de l'enseignant).

Aspects techniques

Le site *Webfinder* est développé avec l'arborescence suivante :

- [css] (*contient les feuilles de style*)
 - jquery-ui.css
 - wfstyle.css
- [db] (*contient l'installation de la table*)
 - change_connect.php
 - install_tables.php
 - webfinder.sql
 - wf_connect.php
 - _wf_connect_data.php (le 1er *underscore* disparaît après installation)
- [img] (*contient toutes les images*)
 - [en] (*les vignettes des pages en anglais*)
 - page01.png
 - page02.png

- page03.png
 - page04.png
 - page05.png
 - page06.png
 - page07.png
 - page08.png
 - page09.png
 - page10.png
- [fr] (*les vignettes des pages en français*)
 - page01.png
 - page02.png
 - page03.png
 - page04.png
 - page05.png
 - page06.png
 - page07.png
 - page08.png
 - page09.png
 - page10.png
- asc.gif (*trois icônes de tri du tableau*)
 - bg.gif
 - desc.gif
 - de.png (*flag*)
 - en.png (*flag*)
 - es.png (*flag*)
 - fin.png (*icône de fin*)
 - fr.png (*flag*)
 - nav_continuer.png (*nav* concerne la navigation*)
 - nav_loupe.png
 - nav_revenir.png
 - shift.png (*icône de la touche shift, pour aide*)
 - wf_header_en.png (*trois icônes pour identifier webfinder*)
 - wf_header_fr.png
 - wf_logo.png
- [include]
 - bdd0.php (*bdd* concerne le tableau de la base de données*)
 - bdd1.php
 - bdd2.php
 - bdd3.php
 - checkalgo.php (*check* concerne les scripts de vérification des réponses*)
 - checkcrit1.php
 - checkcrit2.php
 - checkcrit3.php
 - checkindex.php
 - checkoubli.php
 - checkpub.php
 - form_off.php (*form* concerne la barre de recherche, activée ou désactivée*)
 - form_on.php
 - head.php

- lang.php
- options.php
- pub.php
- row1.php (*row* concerne la colonne de résultats*)
- row2.php
- row3.php
- section_centre.php
- suggestions.php
- tablesorter.php (*pour le tri des colonnes de la base de données*)

- [js] (*contient les scripts de triage de colonnes d'un tableau*)
 - jquery.tablesorter.js
 - jquery-1.12.4.js
 - jquery-latest.js
 - jquery-ui.js

- [lang]
 - en.php
 - fr.php

- erreur.php (*si la recherche ne correspond pas à "dig"*)
- fin.php (*page de fin avec les trois recherches*)
- index.php
- install.php
- intro.php
- moteur0.php
- moteur1.php
- moteur2.php
- moteur3.php
- moteur4.php
- recap.php
- resultats.php
- results.php

Il suffit de copier l'ensemble des fichiers dans un dossier spécifique sur un hébergement et d'avoir les éléments nécessaires à une installation dans une base de données (serveur, utilisateur, mot de passe et nom de la base), afin de procéder à l'installation.

[>](https://profdoc.iddocs.fr/IMG/zip/webfinder2.5.zip "Zip - 1.2 Mo")

Téléchargement de Webfinder 2.5

Les dix pages web de chaque langue sont stockées dans des table (*webfinder_sites_en* et *webfinder_sites_fr*) de base de données, avec la structure suivante :

- id
- titre
- url
- texte
- nb_liens
- nb_visites
- nb_dig
- score_mot

- page_rank1 (*calcul préétabli pour la 1re recherche*)
- score_visites
- page_rank2 (*calcul préétabli pour la 2e recherche*)
- score_liens
- page_rank3 (*calcul préétabli pour la 3e recherche*)
- droit

Les liens commerciaux sont enregistrés dans la table *webfinder_pubs* avec la structure suivante :

- id
- titre
- url
- url_short
- texte
- lang

Les langues sont gérées avec la table *webfinder_langs* avec la structure :

- id
- language
- charset

Une simulation est forcément limitée. Ici, quand on demande aux élèves de faire une recherche autour de la digestion, ceux-ci sont amenés à taper une requête commençant par "dig", qui correspond pour nos pages aux mots du lemme "digestion" (avec des termes qui apparaissent en suggestions).

Toute amélioration est la bienvenue, autant au niveau technique qu'au niveau théorique.

Mises à jour :

- Version 2.5. le 23/06/2018. Correction d'affichage de consignes (partie sur l'indexation).
- Version 2.4. le 31/03/2018. Modification de styles (couleurs, arrondis). Suppression de http:// dans le tableau de la base de données. Mention que les données numériques sont fausses. Amélioration de consignes. Correction d'une vérification de réponse (version anglaise). Correction pour meilleure lecture de réponse à l'addition pour *page rank* (vérification de chaînes multiples).
- Version 2.3. le 10/02/2018. Changements dans les consignes pour aider la compréhension des élèves. Désactivation de la barre de recherche quand elle n'est pas utile dans la progression.
- Version 2.2. le 12/10/2017. Affichage de la solution n°1 pour le mode débutant. Affichage de la solution sur les publicités.
- Version 2.1. le 19/07/2017. Correction de la version anglaise. Correction de la partie de rédaction de définitions.
- Version 2.0. le 02/07/2017. Ajout de parties sur les algorithmes, le droit à l'oubli, l'indexation, la publicité. Création de trois niveaux (débutant, intermédiaire, avancé). Création de trois modes (en classe 1 ou 2 heures, autonomie). Ajout d'une page récapitulative de construction de définitions. Gestion de plusieurs langues, avec création d'une version française et d'une version anglaise.
- Version 1.6. le 19/02/2017. Changement de licence (AGPL). Résolution d'un bug d'accès sans installation. Gestion de mises à jour. Amélioration de l'affichage. Amélioration des interactions (en cas d'erreur de recherche, sur certains aspects au regard des observations en séance). Intégration des réponses dans l'interface. Intégration de nouvelles notions dans les textes explicatifs (pertinence, popularité, notoriété).
- Version 1.5. le 22/04/2016. Correction pour l'installation.
- Version 1.4. le 27/03/2016. Correction pour l'installation.

Comprendre le moteur de recherche avec une simulation

- Version 1.3. le 29/02/2016. Module d'installation de la table dans la base de données.
- Version 1.2. le 20/02/2016. Correction du "Nb de liens externes" par le "Nombre de liens, sur le web, vers cette page". Correction de "Nb" par "Nombre". Reprise de la page récapitulative.
- Version 1.1. le 18/02/2016. Intégration de la notion de "lemme" dans l'affichage de la base de données (lemme "digestion" regroupant tous les dérivés du terme).

Propositions en attente :

- Traduction en allemand.
- Traduction en espagnol.

[1] ROUSSEAU Julien. Webfinder : le nez dans le moteur. *In* Les trois couronnes [en ligne], 2017. Disponible sur : <http://lestroiscouronnes.esmerée.fr/actu/webfinder-le-nez-dans-le-moteur>