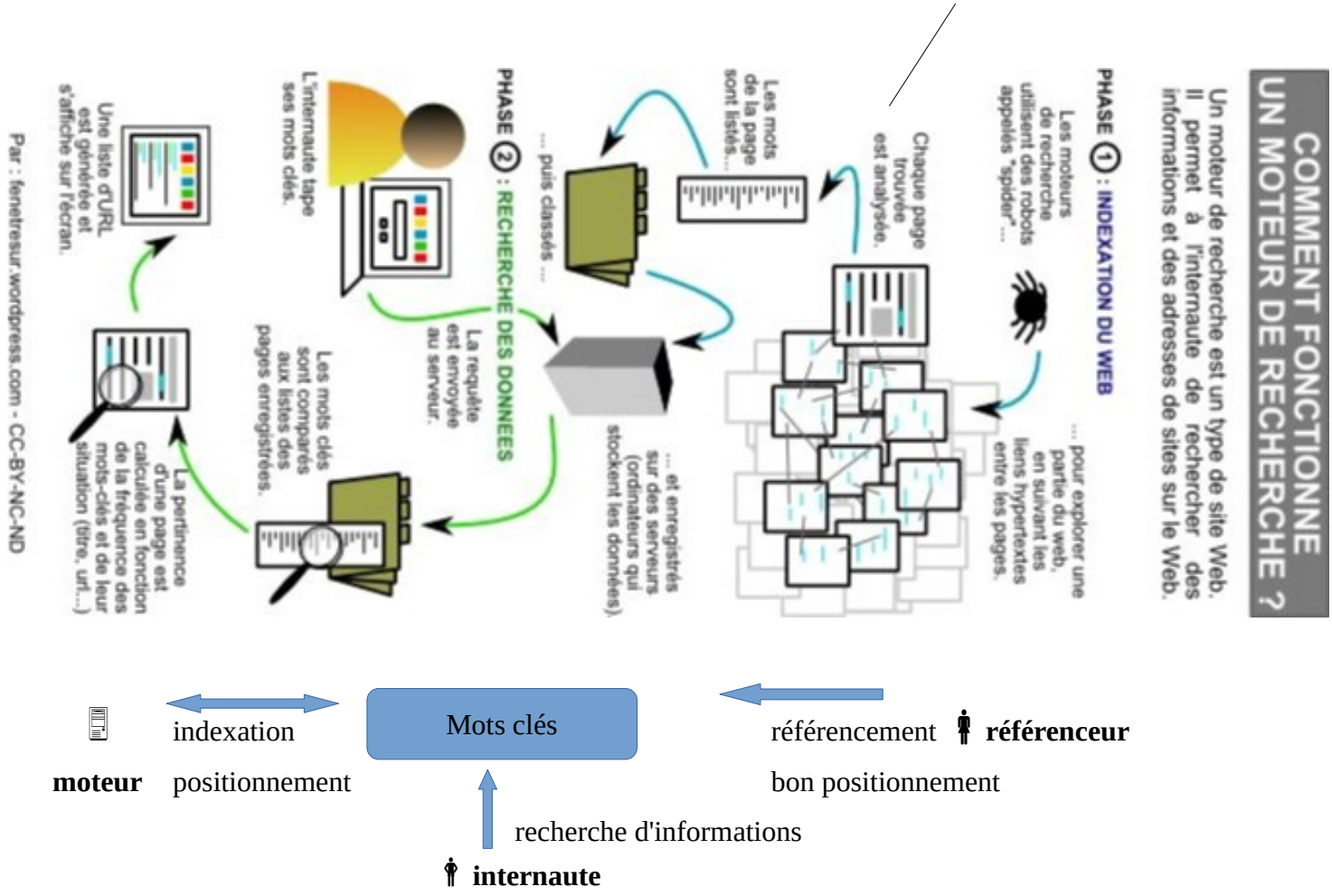
Lorsqu'on cherche une information sur le Web, soit on connaît par cœur l'adresse d'un site, soit on utilise un annuaire (qui regroupe des sites Web), soit on va sur un moteur de recherche. Autant dire que c'est cette dernière solution qui est largement utilisée.
En voici plusieurs :

.....	.....		.....	.....	.....		

Une histoire de mots-clés

Un moteur de recherche fonctionne à partir de mots-clés. Il attend qu'on lui donne des mots clés et il fournit en retour un document hypertexte listant des sites correspondants aux mots clés.

La société qui possède le moteur de recherche utilise en permanence des robots (clients HTTP) qui visitent les sites et suivent les liens qui s'y trouvent. Les plus connus de ces robots sont ceux de Google : les *googlebots*. Ces programmes sont ce qu'on appelle des **robots d'indexation** (ou web crawlers ou spiders) : ils permettent de créer une vaste base de données dans laquelle chaque page Web visitée est stockée et classée.



Sur le moteur de recherche, l'internaute saisit une requête, à l'aide de mots-clés, qui parvient au serveur Web du moteur de recherche. Celui-ci transmet la requête au serveur index : on sait alors quelles pages contiennent les mots clés. Ce serveur index transmet la requête au serveur documents : des snippets (morceaux de code) sont générés en tant que description des résultats. Le serveur documents envoie les résultats à l'utilisateur.

Pour un site, être référencé, c’est bien...

* Faire une recherche en tapant « snt2de » :

- sur Qwant, le site snt2de.glitch.me apparaît-il ? …… - sur Ecosia, apparaît-il ? ……

- sur Google, ce site snt2de.glitch.me apparaît-il ? ……...

Certaines pages Web ne sont pas visitées par les robots donc pas référencées : « ……... »
- les pages statiques : dont le contenu n'évolue pas,
- les pages orphelines : sans liens vers d’autres pages,
- les pages dont l'URL est trop complexe,
- les pages nouvelles,
- les pages qui nécessitent trop de clic pour les atteindre.

Parmi celles qui sont visitées, certaines ne sont finalement pas sauvegardées car provenant de sites illégaux ou de mauvaise réputation (on dit alors « blacklistées »).

… Mais, être bien positionné, c’est mieux !

Pour cela, certains sites payent les moteurs de recherche, on parle alors de référencement payant, par opposition au référencement naturel.

* Avec le mot-clé « nucléaire », relier les sites suivants aux moteurs de recherche si ceux-ci les proposent sur leur première page :

- observatoire-du-nucleaire.org	●	● proposé par Qwant sur sa première page
- voix-du-nucleaire.org	●	● proposé par Ecosia sur sa première page
- edf.fr	●	● proposé par Google sur sa première page

Sur quels critères une page est-elle bien classée ?
Il y a 2 types de critères :

les critères « in page »

Il s’agit du contenu de la page et donc la présence de mots-clés (ou de synonymes).
Ceux-ci peuvent être placés à des endroits stratégiques :
- dans l’URL,
- dans l’en-tête de la page HTML, notamment dans la balise <title>,
- dans les titres et sous-titres (balises <h1>, <h2>)
- dans les informations liées aux images (attribut alt),
- dans les textes des liens..

les critères « off page »

- la popularité des pages : leur fréquentation (nombre de clics),
- la réputation des pages, c’est-à-dire le nombre et la qualité des liens.

* Faire une recherche en tapant « concessionnaire Peugeot » :

- sur Qwant, le garage Sourget de Rennes apparaît-il dans la première page des résultats ? ……...

- sur Google, ce concessionnaire Sourget apparaît-il sur la première page ?……

Google base désormais aussi ses résultats selon la localité du visiteur et l'historique des précédentes recherches effectuées par l'internaute.

Quel avantage y a-t-il à cela ? ……

Quel risque ? ……..

PageRank : la clé du succès de Google

La société Google a été créée en 1998, à une époque où il y avait déjà plus d'un million de sites disponibles. Elle a rapidement écrasé ses concurrents grâce à la rapidité et surtout à la pertinence de ses résultats.

Avec quel pourcentage ? …… réponse d'après https://gs.statcounter.com/search-engine-market-share

Comment ? Notamment grâce à un algorithme créé par Larry Page, cofondateur de Google : le PageRank.

Les pages Web obtiennent une note. Plus une page semble importante, plus elle apparaît tôt dans la liste. La note finale est comprise entre 0 au minimum et 1 au maximum.

Le principe tient en quatre points résumés par l'expression du « surfeur aléatoire » :

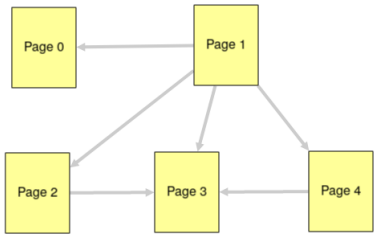
1- il prend une page au hasard, il regarde les liens qu'elle fournit et il suit l'un des liens : « surfeur aléatoire »,
2- plus on détecte de liens vers un site, plus sa note augmente,
3- plus le lien provient d'un site important, plus la note augmente,
4- plus un site crée de liens, moins on donne d'importance aux liens qu'il fournit (cela permet de lutter contre les sociétés tentant d'augmenter artificiellement les notes de leurs clients).

Après la théorie, passons à la pratique. Voici un site proposant une simulation de PageRank **:**

http://computerscience.chemeketa.edu/cs160Reader/_static/pageRankApp/index.html

1ère situation :

On va simuler la situation suivante :
- 5 pages appelées page0, page1, page2, page3 et page4,
- la page1 contient un lien vers chacune des autres pages,
- la page2 contient un lien vers la page3,
- la page4 contient un lien vers la page3.



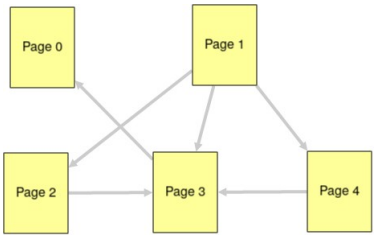
Pour ajouter un lien, cliquer sur le bouton [Add Page] = ajouter une page.
Pour ajouter un lien, cliquer sur la page contenant le lien puis sur la page destinataire.

A votre avis, quelle est la page la mieux notée ? …….. et la moins bien notée ? …… .

Vérifier votre réponse à l'aide du bouton [Run Page Rank]

2ème situation :

Modifier le lien vers la page0 :
ce n'est plus la page1 (mal notée) mais la page3 (bien notée) qui la propose.
Pour supprimer un lien, cliquer dessus puis sur le bouton [Delete Selected]



Est-ce important d'être cité par une page bien notée ? …….

Pour aller plus loin…
On peut établir des statistiques sur la fréquence des mots clés saisis par les utilisateurs et en déduire certaines informations… Aller sur https://trends.google.fr/trends/explore?q=Assomption&geo=FR
Cette page vous donne la fréquence de la recherche « Assomption » sur Google en France.
A l'aide du graphique affiché, deviner quand a lieu la fête catholique de l'Assomption ? ……..
Et quelles « régions » sont sans doute les plus catholiques en France ? …….

Avec Google, on est traqué !