

## Prérequis

Cette étape est valable pour l'ensemble des prochains TP et du projet.

Installer cygwin (<https://cygwin.com/install.html>) en incluant les packages suivants :

- make (The GNU version of 'make' utility)
- gcc-g++ (GNU Compiler Collection C++)
- flex (A fast lexical analyzer generator)
- bison (GNU yacc-compatible parser generator)
- grep (search for regular expression matches in text files)
- curl (Multi-protocol file transfer tool)
- Et tout autre package qui pourrait vous intéresser

(Il faut l'installer en choisissant l'option « pour tous les utilisateurs », et vérifier que les commandes fonctionnent bien sur l'invité de commande. A défaut, ajouter le répertoire des commandes aux variables d'environnement (Ajouter c:\cygwin64\bin à la variable PATH)).

**Alternative :** Ce TP peut aussi être l'occasion de tester le nouveau WSL (Windows Subsystem for Linux) et qui facilite l'installation et l'utilisation des outils cités précédemment sur Windows.

<https://docs.microsoft.com/en-us/windows/wsl/>

Vous pouvez aussi vous entraîner sur les sites :

- <https://regexr.com/>
- <https://www.regexpal.com>
- <https://regex101.com/>
- <https://jex.im/regex/>

## Exercice 1 : grep

En utilisant la commande «grep», trouvez dans le fichier texte (à télécharger):

- 3 urls,
- une adresse IP,
- une adresse mail,
- une date,
- une couleur,
- un numéro de téléphone
- et un mot en français.

Il faut que votre commande ne renvoie que les réponses correctes (pas de faux positifs supplémentaires).  
Jouez le jeu : Ne trichez pas en lisant le contenu du fichier.

## Exercice 2 : regex en C++

La référence pour la bibliothèque STL <regex> de C++ est disponible sur la page :

<http://www.cplusplus.com/reference/regex/>

Ecrire un programme C++ permettant de vérifier si une chaîne de caractère correspond à une expression régulière.

Tester ce programme pour saisir des informations correctes à partir d'une entrée standard (std::cin). Le programme doit boucler tant que l'information saisie n'est pas conforme :

- Un nom d'utilisateur (caractères autorisés, longueur, etc.)
- Un mot de passe « fort »
- Une adresse email
- ...

Essayez les expressions régulières discutées ici :

<https://digitalfortress.tech/tricks/top-15-commonly-used-regex>

## Exercice 3 : Curl et web scraping

Utiliser la commande « curl » pour récupérer le contenu HTML d'une page

A partir de ce contenu et en utilisant le mécanisme d'acheminement des entrées sorties, combiner « curl » et « grep » pour :

- Récupérer le titre de la page
- Récupérer l'ensemble des liens qu'elle contient (balise href= « ... »)
- Récupérer l'ensemble des images qu'elle contient
- Récupérer les métadonnées de la page