Projet de Fin d'Etudes – Promo 2019
AcademyIn : Big Data et Systèmes de Recommandation

Mohamed Quafafou
Mohamed.quafafou@univ-amu.fr

Option : InSI

Description :
La recherche et la géo-localisation d'experts est un enjeu important pour la société moderne d'information et de communication. Comment alors recommander un expert pour une personne qui effectue une recherche de compétences particulières. Plus généralement, ce problème devient important puisque les experts sont recherchés partout dans le monde. Plus généralisation, les systèmes de recommandation deviennent importants pour le développement de la société d'information et de communication.

Figure 1. Utilité de l'expert

Mots clé : Big Data, Systèmes de recommandation, Académie,

Problématique :
Nous allons nous focalisons dans ce projet sur les experts universitaires ou académiques en utilisant des données collectées et concernent leurs publications scientifiques. De tels documents contiennent plusieurs informations entre autres les noms des universitaires, les personnes avec qui ils collaborent (les co-auteurs), leur affiliation, des informations sur leurs activités scientifiques, dans quels évènements ils participent, etc.

Tâches :

Les principales tâches sont :

1) *Etude fonctionnelle pour expliciter les principales fonctions à mettre en place. Puis, définir une architecture de l'application et spécifications des interfaces*
2) *Migration des données et Stockage NoSQL des données en utilisant MongoDB et Neo4j*
3) *Développement de fonctions de base : gestion de profile, connexion, etc.*
4) *Intégrer des API de « text Mining »*
5) *Développement de plugin (pour navigateurs) de collecte de données, etc.*
6) *Développement de fonction de recommandations*
7) *Teste et Evaluations en considérant des membres de Aix-Marseille Université*

D'autres informations (utiles) :
Un accompagnement sera assuré pour expliquer la sémantique des données, et vous aider à spécifier les principales fonctions à implémenter. Pour vous faciliter le développement, vous on vous suggère d'utiliser AdminLTE.

Projet de Fin d'Etudes – Promo 2019
BigDP : Big Data Plateforme

Mohamed Quafafou
{prenom.nom@univ-amu.fr}

Option : InSI

Description :

Actuellement, le domaine de l'informatique manque des outils qui traitent les données massives automatiquement, vue que le volume des données générés par les systèmes d'information des entreprises est devenu colossal, dans ce sens il existe une exigence réelle d'une solution pour gérer la complexité des traitements de ces données. En effet, les entreprises sont de plus en plus nombreuses à prendre conscience de la dimension stratégique de l'information. Ainsi, il y a encore quelques années, lorsqu'il était question de la gestion des données au sein d'un organisme. Désormais, après l'apparition de l'internet, le Volume, la Variété, la Vélocité et la Véracité des données sont devenues d'une grande envergure. Aujourd'hui le problème n'est plus d'avoir accès à l'information mais de la sélectionner et de trouver la bonne information au bon moment.

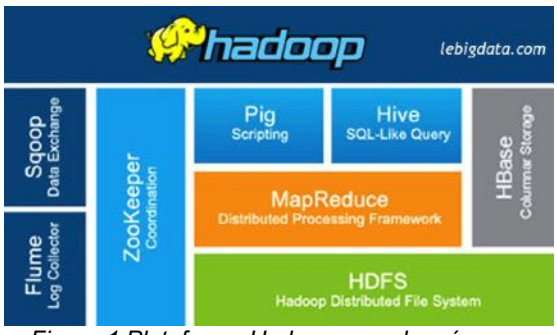


Figure 1.Plateforme Hadoop pour données massives

Mots clé : Big Data, Plateforme, Open source, Configuration, Sécurité

Problématique :

L'objectif est la mise en place et la réalisation d'une application desktop paramétrable qui va offrir un point d'entrée unique qui accepte de différentes sources de données, afin que ces dernières soient pré-exploitable en sortie par d'autres applications.
Pour répondre à la problématique, notre sujet a pour objectif de réaliser une solution en se basant sur une multitude de Frameworks principalement Hadoop, Spark, Flink, Flume, Hive, Mahoot, Sqoop, qui vont permettre depuis une diversité de sources de données, l'extraction, la transformation, et le chargement des données vers une base de données uniforme prête à l'exploitation. Avec en plus, une IHM cohérente, fonctionnelle et qui a pour but de faciliter le paramétrage et la configuration de l'ensemble des briques qui constituent la solution.

Tâches :

Les principales tâches sont :

1) *Elaboration du cahier des charges afin de préciser et de mettre en évidence les principaux axes sur lesquels notre projet va être construit.*
2) *Formation et préparation de l'environnement, cela comprend aussi la familiarisation avec les outils et les technologies qu'on va utiliser au cours de la réalisation du projet.*
3) *La réalisation de la couche applicative qui va jouer le rôle d'un conteneur qui structure les différentes briques du système.*
4) *Tests finaux des différents modules du projet et finalisation du livrable.*

D'autres informations (utiles) :

Une première plateforme a été déjà constituée et vous allez commencer par son l'étude via sa documentation technique, sa mise en place en installant et configurant chacun de ses composants, et son teste sur une infrastructure matérielle distribuée. Ensuite, vous aller la critiquer pour en apporter d'autres améliorations et développer une application faisant la preuve du concept.

Projet de Fin d'Etudes – Promo 2019
BiGraph : Performances des Systèmes Grands Graphes

Malek Habi, Option : InSI
{prenom.nom}@univ-amu.fr

Description :
Les graphes sont des modèles mathématiques puissants capables à la fois de capturer les différentes relations entre les données, leur donner une meilleure représentation et de faciliter leur exploration. Ils sont utilisés dans une large gamme d'applications réelles (Facebook, Google, Twitter, LinkedIn…).
L'objectif de ce projet est d'effectuer une étude comparative entre quelques systèmes existants (Giraph, GraphLab, Neo4j distribué, etc.) pour mieux se familiariser avec la distribution de données et de construire une application dédiée à la vie quotidienne des étudiants en utilisant toutes les connaissances requises.

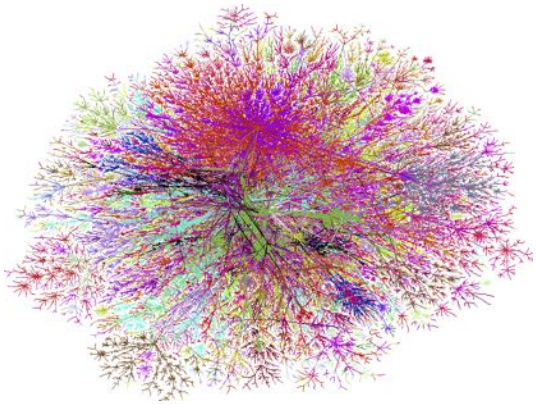


Figure 1.Grand Graphe

Mots clé : Big Data, Systèmes de gestion de grands graphes, Réseaux sociaux, *….*

Problématique :
Les contraintes résultant des données massives (Big Data) se traduisent par un changement brutal de paradigme dans le domaine informatique. En effet, les systèmes classiques ne peuvent fonctionner correctement en respectant ses contraintes. Des nouvelles plateformes avec des capacités de stockage et de traitements importantes, ont vu le jour. Elles sont nécessaires pour indexer, gérer, explorer et analyser ces grandes masses de données.
Néanmoins, les caractéristiques et la diversité de ces nombreux systèmes posent des difficultés pour la plupart des utilisateurs et des chercheurs. Le manque de connaissances des forces et des limites de chaque système rend leur comparaison très difficile. Cela empêche les utilisateurs de décider du meilleur système pour leurs applications.

En plus de se familiariser avec les systèmes de gestion de grands graphes distribués, l'objectif de ce projet est double : (1) effectuer une étude comparative entre différents systèmes (Giraph, GraphLab et Neo4j distribué) et (2) développer l'application LinkedAc dédiée aux élèves des écoles du réseau Polytech.

Tâches :
Description des tâches à réaliser

Les principales tâches sont :

1) Tâche 1
2) Tâche 2
3) Tâche 3
4) Tâche 4
5) …

D'autres informations (utiles) :
Portail existant, documentation complète concernant son analyse, sa conception, son développement et son déploiement.

Projet de Fin d'Etudes – Promo 2019
Coopération interactive 3D en réalité virtuelle

Marc DANIEL, Sébastien MAVROMATIS
{prenom.nom@univ-amu.fr}

Option : ReVA

Description :

Coopération et collaboration sont deux mots populaires du domaine de la réalité virtuelle. La manipulation coopérative peut-être définie comme une situation dans laquelle deux utilisateurs ou plus interagissent sur le même objet de manière simultanée mais coopérative.

L'objectif de ce projet est de développer un prototype dont le but est d'expérimenter la coopération autour d'une surface 3D.

Un premier développement a été initié autour de l'interaction avec une surface 3D en environnement virtuel. Ce travail peut servir de point de départ au développement de ce nouveau prototype.

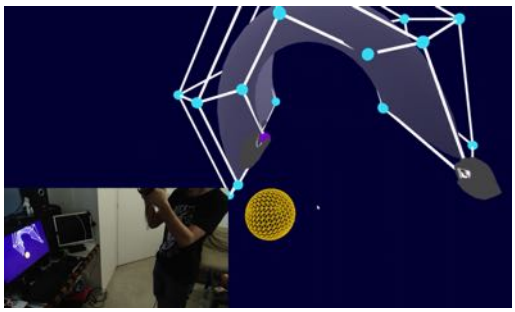


Figure 1. Manipulation d'une surface en RV

Mots clés : Réalité virtuelle, Interaction 3D, Coopération

Problématique :

Comme il a été vu en cours, la qualité perçue d'une surface dépend du point de vue.
Il s'agira d'expérimenter différents modes de collaboration / interaction / visualisation :

- *On pourrait donc imaginer que l'utilisateur 1 demande à l'utilisateur 2 d'évaluer une modification (coopération)*
- *On pourrait aussi imaginer que deux utilisateurs travaillent simultanément (collaboration). Les actions de l'un influencent éventuellement celles de l'autre.*
- *On pourrait s'interroger sur la présence de « pseudo-avatar » des acteurs*
- *On pourrait envisager de travailler avec des périphériques différents (Touch, Wiimote, Leap …)*

Une réflexion sur l'architecture matérielle et logicielle à mettre en place pour envisager un usage distant du prototype devra être menée.
Technologies : Unity 3D, Oculus SDK

Ce sujet est proposé pour un groupe de 4 élèves.

Tâches :

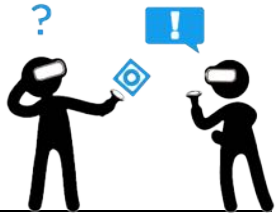
Les principales tâches sont :

1) Prise en main du code existant
2) Amélioration du rendu 3D
3) Développement d'une interface permettant la communication entre deux (ou plus) « sessions » de l'application
4) Adaptation de la vue VR à l'aspect coopératif (Comment signaler les actions entre les différents acteurs ? Comment représenter les acteurs « présents » dans l'environnement ? …)

D'autres informations (utiles) :

Différentes ressources sont mises à disposition :

- *Code source du prototype de manipulation de surface en RV*
- *Support de présentation du travail précédent (rapport, présentation)*
- *Articles scientifiques sur le sujet*



Projet de Fin d'Etudes – Promo 2019
Crypto-monnaie Data Analytics

QUAFAFOU Mohamed
mohamed.quafafou@univ-amu.fr Option : InSI

2 GROUPES

Description :

Une crypto-monnaie ou monnaie virtuelle, est un jeton échangeable entre particuliers sur le réseau Internet. De manière générale, le statut juridique des crypto-monnaies varie considérablement d'un pays à l'autre. Pour certains États, les crypto-monnaies, ou certains d'entre eux sont légalement reconnus comme moyen de commerce, dans d'autres le statut n'est pas encore défini alors qu'enfin la législation concernant les crypto-monnaies est encore en train d'évoluer ou les interdit totalement.

Figure 1. Quelques Crypto-monnaies

Mots clé : *Crypto-monnaie, Data mining, Prédiction, Quotation,*

Problématique :

Il n'y a pas une crypto-monnaie, mais il y en a plusieurs dans les valeurs changent continuellement comme pour les monnaies réelles. Les plateformes de trading spécialisées permettent « de savoir » dans quelle devise investir. Pour cela, de telles plateforme utilise des systèmes qui permettent de prédire la valeur, et donc l'évolution à la hausse ou à la baisse, d'une monnaie. Le but de ce PFE est d'étudier le problème de trading des crypto-monnaies.

Pour répondre à ce problème, vous allez vous proposer un processus de data mining en commençant par l'analyse de l'information reliés à l'évolution des valeurs de crypto-monnaies. Cela vous permettra d'avoir une représentation, structurée et/ou semi-structurée, de vos données. Vous pouvez commencer par le cas de Bitcoin, puis généraliser aux autres crypto-monnaies. En respectant cette représentation et allez collecter les données sur les valeurs de crypto-monnaies à partir de source de données sur le web.

Une fois la collecte de données d'entrainement est collectées différents algorithmes d'apprentissage seront appliqués pour effectuer différentes tâches : regroupement de de crypto-monnaies similaires, prédiction de valeurs, expliciter les indépendances (s'elles existent) entre crypto-monnaies, etc. Le processus de data mining utilisera une plateforme de big data que vous allez mettre en place en utilisant différent outils open source.

Tâches :

Groupe 1 *– Crypto-monnaie DATA :*

1) Etude de crypto-monnaies et proposition d'une représentation (indicateurs pertinents)
2) Etudes de sources de données sur le web qui fournissent des informations sur les crypto-monnaies,
3) Collecte des données du web
4) Evaluation d'algorithmes pour la prédiction de quottes de crypto-monnaies

Groupe 2 *– Crypto-monnaie PLATEFORME :*

1) Collecte de Données (FLUME, …)
2) Programmation (Spark MLlib, streaming et SQL, etc.)
3) Interface utilisateur (Hue)
4) Sécurité : Knox, et bien d'autres outils réaliser d'autres taches.

Autres :

Quelques références sur le web seront mises à votre disposition pour guider vos actions.

Projet de Fin d'Etudes – Promo 2019
Data Quality and Machine Learning : A Medical Case Study

Agus Budi Raharjo, Mohamed Quafafou
agus-budi.RAHARJO@univ-amu.fr, mohamed.quafafou@univ-amu.fr

Option : InSI

Description :

Le cancer de la peau (mélanome) représente le plus important cancer chez l'être humain, et son nombre ne cesse d'augmenter avec le temps. Parmi les nombreux types de cancers de la peau, le mélanome reste aujourd'hui l'un des plus graves, puisqu'il entraine la mort dans 75% des cas. Ainsi, de nombreux chercheurs ont étudié ce problème afin d'arriver à une solution pour détecter ce cancer le plus tôt possible chez les patients [2,3,4].

Figure 1. Cycle de vie

Mots clé : *….*

Problématique :

Nous allons alors développer une application médicale réelle, Mélanome, dont le but est de prédire le potentiel d'un mélanome à partir d'une image. En outre, nous prédisons le niveau de difficulté pour prédire l'image avant son envoi à la dermatologie. Ce niveau de difficulté sera utile pour informer que l'image est prise correctement ou non.

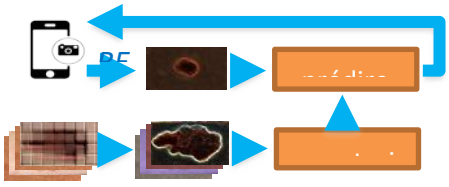


Figure 1. Organigramme de la méthode proposée.

Tâches :

Les principales tâches sont :

1) Frontal (front-end)
    a. Acquisition d'images de mélanome (applications mobiles)
    b. Traitement d'image pour représenter le mélanome (sur le serveur)
    c. Prédire la qualité et la classe de l'image en utilisant les services web exposés (voir arrière-plan)
    d. Utiliser AdminLTE[1] pour le développement d'une application inter-appareils (ordinateur, mobile, PDA)
2) Arrière-plan (back-end)
    a. Utiliser une méthode Apprentissage Automatique actuelle (en tant que service Web Restful)
    b. Utiliser un jeu de données d'apprentissage pour construire un modèle
    c. Utiliser une méthode pour prédire la qualité et la classe d'une image

D'autres informations (utiles) :
Portail existant, documentation complète concernant son analyse, sa conception, son développement et son déploiement.

1. https://adminlte.io/themes/AdminLTE/index2.html
2. Ballerini, L., Fisher, R. B., Aldridge, B., and Rees, J. (2012). Nonmelanoma skin lesion classification using colour image data in a hierarchical K-NN classifier. In ISBI, pages 358–361. IEEE.
3. Gareau, D., Hennessy, R., and Jacques, S. (2012). Automated detection of melanoma. Google Patents.
4. Pereyra, M. A., Dobigeon, N., Batatia, H., and Tourneret, J.-Y. (2012). Segmentation of skin lesions in 2d and 3d ultrasound images using a spatially coherent generalized rayleigh mixture model. IEEE Transactions on Medical Imaging, 31(8) :1509–1520.

Projet de Fin d'Etudes – Promo 2019
Données du cerveau : de l'acquisition à l'analytics

Mohamed Quafafou
Mohamed.quafafou@univ-amu.fr

Option : InSI

2 Groupes

Description :
Il devient possible de récupérer les données émises par le cerveau en particulier les signaux EEG (électroencéphalogramme). Ces données sont de plus en plus exploitées pour construire des modèles une ou un groupe personnes. Pour cela, il faut d'abord utiliser un système d'acquisition des signaux du cerveau. Nous utilisons le casque Muse que vous avez vu en TD/TP. La technologie Emotiv Epoc+ est un système biofeedback multicanal sans fil, haute résolution tout en étant portable.

Figure 1. Emotiv Epoc+

Mots clé : Cerveau, Data, EEG, e-santé

Problématique :
Nous allons utiliser les deux casque Muse et Epoc+ pour acquérir les signaux du cerveau en étant dans différents contextes (travailler, discuter, étudier, manger, dormir, etc.). Dans un premier temps, on commence par définir un protocole d'acquisition des données. Puis, on étudiera leur stockage en considérant les deux types de casques. Ensuite, on étudie un ensemble d'outils existants pour traiter les données du cerveau, ce qui conduira à une comparaison entre les différents outils en précisant les avantages et les inconvénients pour chacun.
Il faudra aussi pré-visualiser les données, les synthétisés, et les visualiser avec des technologies appropriées.
Se pose alors la question d'interpréter les données obtenues de cerveau en conduisant une expérimentation réelle impliquant au plusieurs dizaines de personnes. Chaque personne est placée dans un contexte et on enregistre les données de son cerveau. Puis, vous effectuer une étude comparative sur les résultats obtenus en appliquant des algorithmes de data mining pour effectuer des regroupements de personnes se comportant de la même manière (clustering), prédire des contextes particuliers en fonction de l'état « émotionnelle » d'une personne (prédiction, classification, etc.).

Tâches :

Les principales tâches sont :

1) *Etude comparaison de chacun des casque*
2) *Acquisition des données*
3) *Stockage, prétraitement, et visualisation*
4) *Etudes des outils dédiés aux traitement des données EEG*
5) *Acquisition des données dans des contextes différents en considérant plusieurs dizaines de personnes*
6) *Application des algorithmes de data mining*
7) *Développement d'applications pour illustration*

D'autres informations (utiles) :
On vous accompagnera dans la définition des objectifs, la compréhension des données, l'utilisation des casques, etc.

Projet de Fin d'Etudes – Promo 2019
Drone VR

Sébastien MAVROMATIS
{prenom.nom@univ-amu.fr}
Option : ReVA

Description :

La législation encadre le pilotage de drone. Actuellement un test en ligne permet d'obtenir un certificat d'aptitude.
L'essor de ce type d'objet volant va certainement faire à nouveau évoluer la législation qui va tendre vers l'obtention d'une licence de pilotage.
Un environnement de pilotage virtuel peut apporter une solution à la formation au pilotage et permettre ainsi aux utilisateurs un outil efficace et ludique pour devenir expert en pilotage !

Figure 1. Centre de formation pour le pilotage

Mots clés : Réalité virtuelle, Interaction 3D, Drone, Formation

Problématique :

Le projet a pour but le développement d'un environnement virtuel pour la formation au pilotage de drone.

L'application pourra proposer différentes fonctionnalités comme :
- *Le changement de la météo*
- *Le pilotage avec ou sans l'assistance des capteurs modélisés sur le drone*
- *Des parcours que l'utilisateur devra suivre*
- *Des intrus dans l'environnement comme des oiseaux par exemple*

Technologies : Unity 3D, Oculus SDK

Ce sujet est proposé pour un groupe de 2 à 4 élèves.

Tâches :

Les principales tâches sont :
1) Modélisation d'un environnement virtuel pour le vol
2) Modélisation d'un objet volant
3) Développement de l'application principale
4) Simulation de capteurs d'obstacle sur l'engin volant
5) Amélioration du comportement de l'objet volant (loi de navigation)
6) *Interface avec une radiocommande DJI*
7) En extension : les conditions climatiques, les intrus …

Projet de Fin d'Etudes – Promo 2019
God Game VR

Sébastien MAVROMATIS
{prenom.nom@univ-amu.fr} Option : ReVA

Description

Inspirée des jeux de stratégies telle qu'Age of empire ou bien Black & White, agrémenté de composantes de jeux de la vie, notre jeu en réalité virtuelle propose de prendre place dans le fauteuil d'une divinité et de voir votre monde évoluer, votre civilisation vous appartient, à vous de l'aider à prospérer ou bien de l'en empêcher si vous le souhaiter.

Pour cela, de nombreux pouvoirs seront à votre disposition, créer une forêt et des terres cultivables pour donner bois et nourritures à vos personnages, ou bien les balayer à coups de météorites ou de tornades.

Figure 1. Planète

Mots clé : Réalité virtuelle, Jeu de la vie, Oculus, Unity

Problématique :

Le projet a pour but le développement d'un jeu vidéo de type « God Game » en réalité virtuelle.

Le « God Game » est un sous-genre du jeu de la vie ou le joueur va pouvoir interagir et influencer directement le déroulement du jeu via différents moyens (pouvoirs, actions …).

Les contrôles VR fourni par l'oculus permettront de pouvoir observer la planète Virtuelle via par exemple rotation, déplacement …

Cela permettra aussi de pouvoir agir sur les entités qui se trouve dessus (personnage, ressources…)

Plusieurs idées peuvent être ajoutées aux mécaniques et règles du jeu selon l'état de l'avancement :

- *Génération aléatoire de la planète*
- *Multiples planètes*
- *Évènement aléatoire*
- *Différents Modes de jeu (mode Dieu versus humain …)*
- *Options de Scores*
- *Ajout d'audio*

Technologies : Unity 3D, Oculus SDK

Tâches :

Les principales tâches sont :

1) Modélisation d'une planète
2) Implémentation d'interaction avec la planète (rotation, gravité planétaire …)
3) Création et modélisation de Personnage
4) Création et modélisation des ressources
5) Ajout d'interactions entre les entités (personnage qui amasse des ressources …)
6) Ajout d'interactions avec le joueur (pouvoirs …)

D'autres informations (utiles) :
Portail existant, documentation complète concernant son analyse, sa conception, son développement et son déploiement.

Projet de Fin d'Etudes – InSI – Promo 2019
Intégration d'applications d'Architecture (BimArt)

Tuteur : A. Boukara (Ecole d'architecture, Luminy)
abdelaziz.boukara@marseillearchi.onmicrosoft.com

Option : InSI

Description :

Comme tous les domaines, l'architecture est entrain d'évoluer en profitant pleinement des apports des nouvelles technologies pour le développement de bâtiments durables et intelligents, Dans ce but, plusieurs domaines de compétences collaborent, coopèrent ou se succèdent. Ils utilisent des outils logiciels de plus en plus puissants. Des échanges de données, de contraintes et de paramètres sont nécessaires et impactent les travaux de chaque intervenant. Malheureusement, le manque d'outils communs et les différences des logiciels utilisés dans chaque corps de métier rend difficiles les échanges
Le BIM (Building Information Modeling) est un processus défini pour favoriser le travail entre les différents intervenants dans un projet de conception et construction. Le challenge aujourd'hui est de créé une plateforme numérique, supportant le processus BIM, visant à faciliter les échanges de données, de résultats, contraintes et communications entre différents outils logiciels (Autocad, Revit,Archicad, Transys, Pléiade Comfie, Matlab, Etc…

Mots clé : *Bâtiment durables, Energies renouvelables, Thermique, Architecture, Design, BIM, Conception, CAO, Simulation.*

Problématique :
La plupart des outils dédiés à la conception et la saisie graphique destinés à l'architecte ne sont ni adaptés ni ouverts à la réutilisabilité. Leurs formats d'échanges et d'exports vers d'autres disciplines et environnements d'ingénierie tel que les Environnements de simulation thermique et énergétique ne permettent pas une souplesse et une exactitude dans la récupération des données et d'informations saisies au préalable, à savoir les données graphique et alphanumériques liés au bâtiment.
La multidisciplinarité de l'acte de construire impose le recours à des environnements divers et complémentaires. La gestion d'une multitude d'informations reste fastidieuse, du fait du nombre de logiciels utilisés, et du nombre de formats de fichiers existants. Les Outils CAO ne permettent pas une grande flexibilité et interopérabilité avec les outils énergétiques. Ils ne sont pas basés sur des approches automatisées,
Ils existent des solutions, telle que les IFC et le Gbxml, mais malheureusement ces approches n'ont pas encore atteint l'objectif souhaité. Nous sommes loin de la compatibilité totale.
Ils existent néanmoins des passerelles de part et d'autre, à savoir dans les deux sens, Architecture et ingénierie. Ces dernières ne sont pas abouties.
Prenant l'exemple d'un outil comme TRNSYS qui intègre dans son package un plugin Sketchup (Trnsys3D) celui-ci est destiné à l'import et la lecture d'un fichier généré par un outil de CAO tel que ArchiCAD ou REVIT :
Une fois importé, ce fichier ne sert qu'à la lecture du contour graphique en plan (fond de calque). De ce fait, toutes les informations liées à la 3D et au renseignement de l'entité graphique sont alors perdues. Le plugin a en effet transformé une entité 3D en une entité 2D, donc plus pauvre en informations qu'a l'état initial. Le thermicien est amené à redessiner le modèle en 3D et le renseigner encore une fois. Le même problème se pose avec les mêmes conséquences au niveau de l'utilisation des outils comme PLEIADES COMFIE, CLIMAWIN, Energy+ etc……

Tâches :
Les principales tâches sont :

1) Se familiariser avec la modélisation d'information du bâtiment (BIM)
2) Etude comparative des outils et logiciels dédiés aux bâtiments
3) Normaliser les échanges de données entre les différents outils
4) Développer une plateforme d'intégration en exploitant les normalisations proposées
    a. Analyse et conception
    b. Choix d'un exemple de bâtiment à modéliser
    c. Exploiter les outils de façon séquentiels en automatisant leur intégration à l'aide de données
    d. Etudier les différents autres modes d'intégration des outils et étudier leur impact sur leur intégration
    e. Tester et validation

D'autres informations (utiles) :
Mise à disposition des outils de modélisation, support et formation en ce qui concerne l'utilisation de certains outils à des fins de modélisation, conception, optimisation, gestion et contrôle de l'énergie ; Initiation à l'architecture

Projet de Fin d'Etudes – Promo 2019
Jeu en réalité augmentée et Recommandation

Mohamed QUAFAFOU, Sébastien MAVROMATIS
{prenom.nom@univ-amu.fr}
Option : InSI - ReVA

Description :

Le domaine de la réalité augmentée (AR) connait un essor important. Ce sujet aborde certains aspects relatifs à l'AR comme le calibrage de caméra, la détection de points d'intérêts, le suivi de points d'intérêts, le rendu d'objets virtuels dans un environnement réel … De plus, les applications d'AR demandent à l'utilisateur l'interprétation d'une scène réelle « enrichie » et l'étude de l'impact de ce type de scène sur la charge cognitive de l'utilisateur est intéressante. Ce sujet propose, à travers un jeu simple, d'appréhender ces deux aspects en proposant un jeu de labyrinthe couplé à un système de coaching pour pousser le joueur hors de ces limites.

Figure 1. Labyrinthes !

Mots clés : Réalité augmentée, Capteurs, Recommandation, Charge cognitive

Problématique :

L'objectif de ce travail est :

- *de proposer une application graphique permettant à l'utilisateur d'interagir par l'intermédiaire d'un objet dont les caractéristiques sont automatiquement extraites du flux vidéo d'une caméra*
- *de mettre en place un système de « coaching » pour le joueur en fonction de données représentatives de l'activité cérébrale*

Ce sujet est proposé pour un groupe de 2 élèves.

Tâches :

Les principales tâches sont :

1) Récupération du flux vidéo de la webcam, capture de la « planche de jeu »
2) Superposition d'un modèle 3D de la planche de jeu sur le flux vidéo
3) Génération du labyrinthe à partir des tracés présents sur la planche de jeu
4) Modélisation des éléments du jeu : la bille, les murs, les trous, le départ, l'arrivée …
5) Mise en place de la physique du jeu, rendu 3D
6) Interfaçage avec les capteurs, récupération des données
7) Mise en place du système de recommandation

Autres :

Un prototype existant sera mis à disposition.

Projet de Fin d'Etudes – Promo 2019
Plate-forme d'analyse et de visualisation de trajectoires

Nicolas DURAND, Mohamed QUAFAFOU
{prenom.nom@univ-amu.fr}
Option : InSI

Description :

L'analyse du trafic routier dans une ville est devenue un enjeu crucial qu'il faut considérer pour pouvoir gérer convenablement une ville en facilitant la vie des citoyens. Ce projet vous permet de vous initier au problème de l'analyse de trajectoires d'objets mobiles (suite de coordonnées GPS) et de la visualisation des résultats de cette analyse.

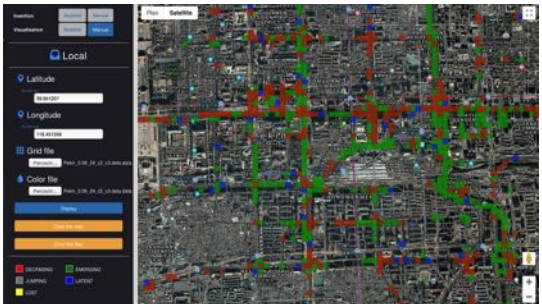


Figure 1. Exemple de résultat d'analyse montrant l'évolution du trafic (avec l'outil existant).

Mots clés : *Data Analytics, Trajectoires, Visualisation, Données Spatio-Temporelles, GPS.*

Problématique :

L'objectif du projet est de développer une plate-forme pour analyser un ensemble de trajectoires d'objets mobiles (par exemple, des taxis) et visualiser les résultats obtenus. Pour cela, le projet s'appuiera sur les programmes existants d'analyse de trajectoires (développé dans le cadre d'une thèse de doctorat) et sur une première version de l'outil de visualisation (résultat de précédents projets).
Pour information, les programmes d'analyse des trajectoires ont été réalisés en JAVA et des scripts Shell ont été utilisés pour automatiser différentes tâches. La partie visualisation a été développée avec Angular js, Node js, MongoDB, GridFS et l'API Google Maps.

Plusieurs points importants seront à aborder, comme le stockage de données volumineuses (les ensembles de trajectoires à analyser), le calcul (l'appel des fonctions d'analyse via une API ou des services), la prise en compte de l'aspect temporel dans la visualisation des résultats.

Au final, la plate-forme devra fournir un portail web opérationnel et accessible à tous les utilisateurs souhaitant analyser ses propres données de trajectoires.

Tâches :

Les principales tâches sont :

1) Evolution de l'outil de visualisation par la prise en compte de tous les paramètres d'analyse.
2) Mise en production du portail web.
3) Développement de l'API ou des services permettant de lancer les calculs d'analyse.
4) Chargement et stockage des données (ensemble de trajectoires) d'un utilisateur.
5) Eventuellement, collecte de trajectoires venant d'utilisateurs volontaires (anonymes).

Autres :

Il sera mis à votre disposition des documents sur le problème de l'analyse de trajectoires (de l'acquisition des données à la visualisation), des services prêts implémentant les principales fonctions liées à l'analyse de trajectoires, les outils développés pour la visualisation et tout un accompagnement pour vous expliquer les concepts de base, les méthodes appliquées, et les autres aspects scientifiques.

Projet de Fin d'Etudes – Promo 2019

Réalisation d'une application graphique pour l'étude du comportement manuel dans une trajectoire restreinte

Peter BANTON, Sébastien MAVROMATIS
{prenom.nom@univ-amu.fr} Option : ReVA

Description :

La Steering Law est une description (modèle mathématique) du comportement manuel dans une trajectoire restreinte. Elle dit que le temps mis pour tracer un chemin d'une forme quelconque est une fonction affine de la "difficulté" du chemin. Pour un chemin de largeur constante (la figure 2), la difficulté (ID) est définie comme :

*ID = Amplitude/(Width*ln(2))*

Cette loi est problématique dans la mesure où elle ne tient aucun compte de la forme du chemin. Elle prédit que l'on mettrait le même temps pour parcourir un chemin droit de longueur D que pour un chemin circulaire de circonférence D. Cette prédiction est fausse : cela a été démontré par des expériences scientifiques.

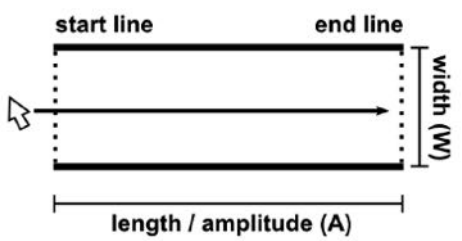


Figure 1. Un chemin droit

Mots clés : Steering law, tablette graphique

Problématique :
La finalité de ce projet est de produire une application qui permet de construire, parcourir et comparer des chemins de formes différentes. La comparaison se fait en chronométrant le temps mis pour parcourir les chemins.

Chaque chemin (droit, courbé) est défini par sa longueur, sa largeur et sa courbure. Tous les chemins sont de largeur constante. La courbure, C, d'un arc de cercle est définie par : C = 1/r, ou r est le rayon de courbature. La courbure d'une droite est donc zéro. La mesure de la courbure totale d'un chemin fait d'éléments droits et courbés est à définir.

Développement :
Ce projet peut être développé soit utilisant une tablette graphique, soit en utilisant une souris :

1. *Créer une application graphique permettant de définir des chemins (sous forme d'image). Différents outils de tracé seront proposés permettant de créer des chemins à partir de caractéristiques géométriques (segment, cercle, ellipse ...) ;*
2. *Eventuellement, créer une application qui s'interface avec une tablette graphique pour parcourir les chemins.*

Les développements seront réalisés en C++/OpenGL.

Tâches :
Les principales tâches sont :

1. Conception/réalisation de la méthode de construction des chemins
2. Calcule de la longueur d'un chemin fait d'éléments différents
3. Calcule de la courbure d'un chemin fait d'éléments différents
4. Conception de l'interface utilisateur
5. Opérationnalisation de la tablette
6. Conception/réalisation de l'expérience (déroulement, chronométrage, traitement des données, etc.)

D'autres informations (utiles) :
https://en.wikipedia.org/wiki/Steering_law

Projet de Fin d'Etudes – Promo 2019
Reconstruction de surfaces à partir de nuages de points

Bac Alexandra (alexandra.bac@univ-amu.fr)
Option : RéVA

Description :

Reconstruction de surfaces à partir de nuages de points 3D (scanners LiDAR)

Participation au développement d'un nouvel algorithme de reconstruction de surfaces à partir de données LiDAR, développé par Jules Morel et Alexandra Bac.

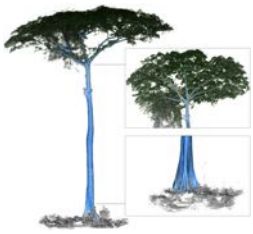


Figure 1. Reconstruction de Poisson sur CSRBF

Mots clé : *nuages de points 3D, LiDAR, reconstruction, surfaces implicites, reconstruction de Poisson, CSRBF*

Problématique :

Un nouvel algorithme de reconstruction de surfaces à partir de nuages de points denses, bruités et largement occlus a été développé dans l'équipe G-MOD par Jules Morel et Alexandra Bac. Il offre un nouveau schéma de Poisson s'appuyant sur un modèle implicite à base de fonctions à base radiale (CSRBF). Mais sa particularité est qu'il permet également d'intégrer des modèles a priori (de type cylindres) pour les zones largement occluses.
Un code prototype a été développé et l'algorithme a été publié l'année dernière. Mais l'objectif est de développer maintenant un code indépendant, optimisé et de nombreuses questions restent à traiter (intégration de modèles plus généraux …).

Tâches :

Les tâches sont essentiellement des tâches de codage et optimisation du prototype existant. Pour cela, il faudra tout d'abord comprendre l'algorithme.

Les principales tâches sont :

1) Compréhension de l'algorithme
2) Etude comparative des bibliothèques PCL / CGAL pour le développement
3) Développement de la reconstruction de Poisson en C++ en s'appuyant sur les bibliothèques Eigen + PCL ou CGAL
4) Développement et optimisation de l'intégration de modèles cylindriques
5) Etude de l'intégration de modèles quelconques

D'autres informations (utiles) :
Plus d'images sur :
https://julesmorel.com
L'article sera fourni au groupe intéressé.

Projet de Fin d'Etudes – Promo 2019
Reconstruction de surfaces à partir de nuages de points

Joris RAVAGLIA, Alexandra BAC
{prenom.nom@univ-amu.fr}
Option : ReVA

Description :

Aujourd'hui les technologies d'acquisition 3D (kineckts, lidars, cameras temps de vol, photogrammétrie, etc) se popularisent. Les instruments manipulés permettent d'acquérir une représentation 3D d'un environnement sous la forme de nuages de points (un ensemble de points dans l'espace). Ces représentations 3D sont ensuite utilisées dans divers domaines tels que : l'urbanisme pour planifier des constructions, la foresterie pour mesurer et évaluer une parcelle forestière, les médias de divertissement pour intégrer des éléments dans des jeux vidéo, films d'animation, et créer des effets spéciaux, etc.

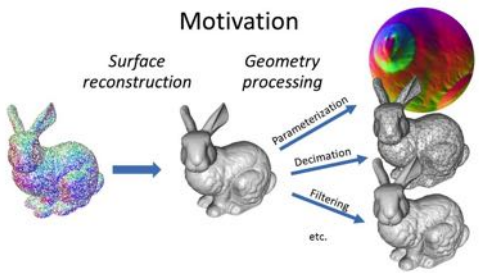


Figure 1. Screened Poisson Surface Reconstruction - Kazhdan and Hoppe

Mots clés : Nuages de points, reconstruction de surface, 3D, LiDAR, maillages

---

Problématique :

Avant d'en arriver à ces applications en bout de chaîne, il est nécessaire de pré-traiter et traiter les nuages de points. Un des points clés dans ces traitements reste la reconstruction de surface à partir du nuage de points : Comment passer d'un ensemble de points éparpillés dans l'espace à une ou plusieurs surfaces géométriques sur lesquelles vont pouvoir être élaborées des procédures plus complexes ?

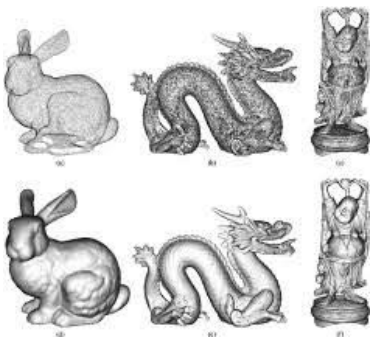
Tâches :

Plusieurs algorithmes ont été mis en place pour arriver à reconstruire les surfaces sous-tendues par un nuage de points. Le but de ce projet sera de :

1. *Faire un rapide tour d'horizon des méthodes existantes*
2. *Comparer la présence ou non de ces méthodes dans des librairies existantes*
3. *Implémenter une méthode de reconstruction particulière inexistante dans les librairies OU comparer les résultats obtenus par différentes méthodes présentes dans une/plusieurs librairies*
4. *Evaluer les résultats obtenus*

Données :

Les données sur lesquelles vous travaillerez seront composées d'un ensemble de nuages de points issus des objets incontournable dans ce domaine (dragon, bouddha, bunny, …) et de données issues du monde forestier.



Compétences :

Durant ce projet, il faudra faire appel à des connaissances en algorithmique, en géométrie, et en c/c++. Vous aurez également à manipuler des logiciels de modélisations tels que meshlab, blender, cloudcompare ou computree.

Projet de Fin d'Etudes – InSI – Promo 2016
Réseau Social des Prof. : cas Polytech Marseille (Profbook)

Mohamed Quafafou
Mohamed.quafafou@univ-amu.fr

Option : InSI

Description :

Les réseaux sociaux sont en plein développement et deviennent un moyen de référence pour la communication et l'échange. Dans la tête de liste on trouve les réseaux[1] *les plus célèbres tels que Twitter, Fracebook, LinkedIn, d'autres qui sont plus spécifiques, par exemple Xing (Allemagne), Renren (Chine), Instagram (photos), Alumnforce (Etudiants et diplômés), et bien d'autres qui sont en émergence. La plupart des réseaux offrent des fonctions pour satisfaire un usage classique tel que le partage rapide de l'information, la connexion aux autres, la publication de messages, etc.*
En plus des fonctions classiques, chaque réseau se différencie par un ensemble d'usages particuliers qui dépendent de la population ciblée, du domaine d'application et bien d'autres paramètres. Cependant, le portail d'un réseau social et son application mobile doivent rester très simples pour être accessibles au plus grand nombre. De plus, les données utilisateurs peuvent être très hétérogènes, volumineuses, avec un taux de mise à jour très élevé, etc. Aussi, les données peuvent avoir des formats multiples représentant d'un simple texte jusqu'à un graphe en passant par les images, les vidéos, les entités nommées, etc., ce qui pose le problème d'optimisation de leur stockage afin de pouvoir les exploiter efficacement. L'autre problème crucial concerne la structuration de l'information et sa visualisation dans le portail et l'application mobile du réseau social. Le but de ce PFE et de vous familiariser avec les réseaux sociaux, en tant qu'informaticien architecte et non un simple utilisateur (clic bouton!). Pour cela, nous allons considérer un réseau social concret à savoir celui des professeurs de l'école Polytech Marseille et plus particulièrement ceux du département informatique.

Mots clé : *Réseau sociaux, Stockage données, Usage, Profile, Web services, Polytech Marseille.*

Problématique :
Comme souligné auparavant, le stockage de données est parmi les problèmes cruciaux qui se posent. Faut-il utiliser des bases de données SQL ou NoSQL ? Oui, Il faut utiliser à la fois les bases de données SQL pour les données structurées (ex MySQL) et NoSQL pour les données non structurées (ex Solr). Plus encore, il faut aussi pouvoir stocker des graphes (ex Giraph) et bien d'autres données. Le second problème concerne la structuration et la gestion de l'information et l'intégration de méthodes dédiées aux réseaux sociaux (par exemple, calcul des communautés). On se focalisera plus particulièrement sur la visualisation de données en utilisant par exemple la technologie d3 (d3js.org).

Tâches :
Les principales tâches sont :

1) Etude comparative de quelques réseaux sociaux[2].
2) Etude et mise en place d'un système de stockage multiple incluant les bases de données relationnelles, NoSQL, Graphe, etc.
3) Développer un réseau social des Professeurs de Polytech Marseille (cas du département Informatique)
    a. Collecter l'information et stockage
    b. Proposer une architecture technique en choisissant les technologies appropriées pour implémenter un réseau efficace, facile d'utilisation et ayant une bonne qualité en ce qui concerne la charte graphique.
    c. Définition et gestion du profil d'un professeur
    d. Développer une solution informatique et l'utiliser pour construire le réseau des professeurs du département informatique. Ce réseau doit-être utilisé via le web, le tel. Mobile et tablette.
    e. Tester et évaluer le réseau ainsi développé

D'autres informations (utiles) :
Données sur les professeurs du département informatique de Polytech Marseille

1. http://www.socialmediatoday.com/social-networks/2015-04-13/worlds-21-most-important-social-media-sites-and-apps-2015
2. http://social-networking-websites-review.toptenreviews.com/

Projet de Fin d'Etudes – Promo 2019
Segmentation et/ou classification de nuages de points

Joris RAVAGLIA, Alexandra BAC
{prenom.nom@univ-amu.fr}
Option : ReVA

Description :

Aujourd'hui les technologies d'acquisition 3D (kineckts, lidars, cameras temps de vol, photogrammétrie, etc) se popularisent. Les instruments manipulés permettent d'acquérir une représentation 3D d'un environnement sous la forme de nuages de points (un ensemble de points dans l'espace). Ces représentations 3D sont ensuite utilisées dans divers domaines tels que : l'urbanisme pour planifier des constructions, la foresterie pour mesurer et évaluer une parcelle forestière, les médias de divertissement pour intégrer des éléments dans des jeux vidéo, films d'animation, et créer des effets spéciaux, etc.

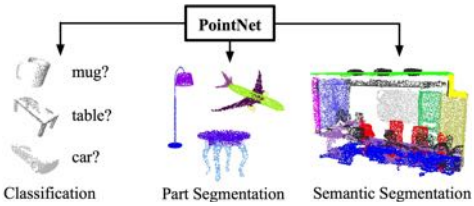


Figure 1. Segmentation de nuages de points

Mots clés : Nuages de points, segmentation, classification

Problématique :

Appréhender un nuage de points brut n'est pas toujours facile… tout comme en traîtement d'image, la segmentation de nuages de points revêt une importance particulière : cette étape est souvent effectuée avant même de reconstruire les surfaces. D'autre part, la classification de nuage de points permet d'analyser finement le contenu d'un nuage de points en vue de reconnaître une forme particulière (ex : avion, mug, etc) ou d'extraire un sous ensemble d'intérêt des données initiales (ex : bois et branches en forêt).

Tâches :

Plusieurs algorithmes ont été mis en place pour segmenter et/ou classifier un nuage de points. Le but de ce projet sera de :

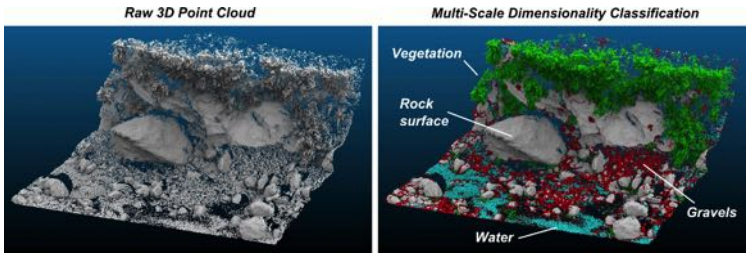
1. *Choisir une méthode de classification ou segmentation*
2. *Tester la validité de cette méthode sur un ensemble de données varié*
3. *Evaluer les résultats obtenus*

Données :

Le jeu de données sera construit au fur et à mesure en fonction de la méthode choisie.

Compétences :

Durant ce projet, il faudra faire appel à des connaissances en algorithmique, en géométrie, et en c/c++. Au besoin, le sujet peut comprendre un réseau de neurones. Vous aurez également à manipuler des logiciels de modélisations tels que meshlab, blender, cloudcompare ou computree.



Projet de Fin d'Etudes – InSI – Promo 2016
Système d'Information ouvert sur le Web (siWeb)

Mohamed Quafafou
Mohamed.quafafou@univ-amu.fr

Option : InSI

Description :
Avec l'avènement de l'économie numérique, les entreprises sont de plus en plus présentes sur le Web, non pas uniquement pour avoir une vitrine digitale, mais pour vendre des produits et services. L'entreprise utilise ainsi le web comme moyen pour son développement en ayant ainsi accès directe à des utilisateurs dont le comportement est volatile et exploite les réseaux sociaux dans leur acte d'achat. Dans ce contexte le système d'information de l'entreprise s'ouvre sur le web dans le sens où il lui fournit les informations nécessaires au développement de ses applications et services en ligne.
Plus encore, le web représente aujourd'hui la plus grande base de données où l'information est généralement représentée sous forme de pages HTML et provenant de millions de sources Web. C'est une base multilingue qui est en perpétuelle évolution et dont le contenu représente des informations liées au monde physique réel dans nous vivons. Ainsi, par exemple, on consulte des sites web spécifiques pour connaitre la météo, les films de la semaine, les résultats sportifs, le nom du directeur de l'école Polytech'Marseille, les programme d'une formation, l'avis sur un produit que je m'apprête à acheter, etc.

Mots clé : *Système d'information, Sources Web, Selenium WebDriver, Automatisation de Tests, Automatisation des Accès, Construction de corpus*

Problématique :
Se pose alors le problème des tests des applications et services en ligne. Qui sont manuels, fastidieux et chronophage : Peut-on automatiser ces tests à l'aide de solutions standardisées ?
L'accès à de une telle information s'effectue manuellement via un navigateur ce qui est aussi chronophage. Comment automatiser l'accès à cette information et sa collecte en vue de construire des corpus de données spécifiques ?
Le but de ce PFE est double : (1) vous faire découvrir une problématique au cœur du travail d'ingénieur d'aujourd'hui à savoir l'automatisation des tests d'applications en ligne, et (2) l'exploitation du web en tant que base de données pour en extraire automatiquement des données spécifiques. Pour atteindre ces objectifs, vous allez considérer des données académique et utiliser la technologie WebDriver [*] qui est un framework permettant l'automatisation dans différents navigateurs tels que FireFox, Chrome, etc. en utilisant différents langages de programmation (Java, C#, Python, PHP, Perl, Ruby). Il est issu du projet Selenium [*].

Tâches :
Les principales tâches sont :

1) Se familiariser avec Selenium WebDirver et les tests d'applications en ligne
2) Exploiter Selenium WebDriver pour automatiser l'accès à l'information web
3) Extraire des données du web concernant les organisations et acteurs du monde académique
    a. Construire une liste de sources web dédiées au monde académiques (Universités, Laboratoires de recherches, etc.)
    b. Exploiter chacune des ressources manuellement afin d'identifier l'information à extraire, repéré les sources les plus fiables, et quantifié la qualité des sources.
    c. Définir ce qui doit-être automatisé et ce qui reste interactif (demandé à l'utilisateur)
    d. Définir une architecture générale simple de la solution permettant d'extraire les informations académiques et de les stocker sur une base de données relationnelle.
    e. Concevoir et développer les modules.
    f. Tester les briques développées et les intégrer dans une seule application en ligne dont il faudra automatiser les tests !

D'autres informations (utiles) :
Initiation au WebDriver ; Code source exemples, liste de source web.

Projet de Fin d'Etudes – Promo 2019
The Wire Loop Game VR

Peter BANTON, Sébastien MAVROMATIS
{prenom.nom@univ-amu.fr} Option : ReVA

Description :

Tout le monde connaît le jeu montré ci-contre (la figure 1). Il consiste en un fil de fer (the wire) autour duquel se trouve une boucle de fil de fer (the loop). L'objet du jeu est de faire passer la loop le long du wire sans le toucher. Chaque fois que la loop touche le wire est une erreur.

Figure 1. Wire loop game

Mots clés : Wire loop game, réalité virtuelle, environnement virtuel

---

Problématique :

Ce projet a pour but la réalisation de ce jeu dans un environnement virtuel.
Il sera demandé de prendre contact avec les enseignants très régulièrement pour valider les choix.

Technologies : ICE ou Unity 3D, Oculus SDK

Tâches :

Les principales tâches sont :

1. Mise en place de l'environnement virtuel ;
2. Création à l'aide d'un fichier de configuration du wire d'une forme variable (longueur, diamètre, sinusoïde, cercle, etc.) ;
3. Création à l'aide d'un fichier de configuration de la loop de diamètre variable ;
4. Création d'un système de détection de position pour détecter les erreurs ;
5. Chronométrage du parcours, sauvegarde des données
6. Affichage du temps de parcours en temps réel dans l'EV

D'autres informations (utiles) :

General Platform for Socially Personalized Speech Synthesis

Project Category: Student's group project proposal v1.1 – October 2018
Acronym: SpSS
Supervisors: NINH Khanh Duy and QUAFAFOU Mohamed
University: Polytech Marseille
Option : InSI, RéVA, InSI-RéVA

ABSTRACT

This project concerns the development of a software platform dedicated for research on text-to-speech (TTS) or speech synthesis in the social context avoiding the classical approach where only one "perfect" speaker is considered [1]. Following the "averaging" approach proposed by Junichi Yamagishi which reduces significantly the amount of data from each speaker [2][3], we take into account the diversity of speakers enriching the usual static and dynamic features with speakers personalized features. Doing so, the learned models encode not only properties of the voice but integrate also personalized features of the speaker.

In order to facilitate socially personalized speech synthesis research, this project is proposed with the aim to develop a general platform for data collection/assessment/validation and dialect analysis. Overall the project focuses around the following problems:

- **Data collection using mobile phones:**
  We try to acquire data (text, speech signal, speaker's profile, etc.) from geo-localized speakers, while not introducing a minimum number. Consequently, we will manage a dynamic multidimensional vocabulary space considering the emergence of new words and phonemes, as well as a large speaker space with various accents for a language. The use of the open-sourced Mozilla's Common Voice project is recommended [4].
- **Speech quality assessment:**
  According to the protocol used to collect speech signals, we need to implement methods to evaluate the quality of speech with high accuracy and reliability. While the listening tests are considered the gold standard in terms of assessment of speech quality, they are costly and time consuming in the context of "big data". For that reason, much research effort has been placed on devising objective measures that correlate highly with subjective rating scores [5][6]. We will put efforts on the implementation of speech quality evaluation methods using objective measures so that the quality of collected speech can be managed automatically and can be used as additional information for the speech data.

- **Pronunciation error detection:**
  Abnormal behaviors during data collection may apear, of which the mismatch between given text and underlying content of collected speech signal (e.g., substitution, deletion, and insertion pronunciation errors) has most severe effect on the performance of speech synthesis system. Although a manual checking step can be done by enrolled speakers through replaying recorded speech, the automatic detection of severe pronunciation errors is still needed when dealing with a large variety of speakers. We will focus on detecting word-level pronunciation errors by leveraging techniques originally proposed for Computer-Aided Language Learning (CALL) systems such as [7][8]. Once such big errors are detected, the corresponding speech signals must be recorded again or removed from the database.
- **Dialect distance measurement and dialect visualization:**
  Once some data has been collected and validated for a language, we can explore fundamental differences between dialects based on available speech data from those dialects. Low-level acoustic and prosodic features of speech signals can be used to study differences between dialects. Dialect proximity measures such as ones proposed in [9][10] can be examined to check whether they are consistent within a language. Once the dialect distance measure is developed, we can display speaker's voices on a "dialect map" according to their dialect distances.

From the practical point of view, our system will collect data from speakers using their mobile phone and display results of our systems (collected signals and their metadata) on a dialect map. At the beginning, we will perform a local experimentation considering speakers from Polytech Marseilles and we will extend it to different locations in France and Vietnam. During this experimental process, we will also evaluate how much our approach is language-independent.

References

[1] K. Tokuda, Y. Nankaku, T. Toda, H. Zen, J. Yamagishi, and K. Oura, *Speech Synthesis Based on Hidden Markov Models*, Proceedings of the IEEE, vol.101, no.5, pp.1234-1252, 2013.

[2] J. Yamagishi, *Average-Voice-Based Speech Synthesis*, PhD thesis, Tokyo Institute of Technonoly, 2007.

[3] J. Yamagishi et al., *Thousands of voices for HMM-based speech synthesis–Analysis and application of TTS systems built on various ASR corpora*, IEEE Transactions on Audio, Speech, and Language Processing, vol.18, no.5, pp.984-1004, 2010.

[4] https://voice.mozilla.org/en/new

[5] Hu, Y., and Loizou, P. C., *Evaluation of objective quality measures for speech enhancement*, IEEE Transactions on Audio, Speech, and Language Processing, vol.16, no.1, pp.229–238, 2008.

[6] Loizou P.C., *Speech Quality Assessment*. In: Lin W., Tao D., Kacprzyk J., Li Z., Izquierdo E., Wang H. (eds) Multimedia Analysis, Processing and Communications. Studies in Computational Intelligence, vol 346. Springer, Berlin, Heidelberg, 2011. (Available online at: https://ecs.utdallas.edu/loizou/cimplants/quality_assessment_chapter.pdf)

[7] A. Lee and J. Glass, *Comparison-based Approach to Mispronunciation Detection*, Proceedings of 2012 IEEE Spoken Language Technology Workshop (SLT), pp. 382-387, 2012.

[8] S. Wei et al., *A new method for mispronunciation detection using Support Vector Machine based on Pronunciation Space Models*, Speech Communication, vol.51, no.10, pp.896–905, 2009.

[9] M. Mehrabani and J. H. L. Hansen, *Automatic analysis of dialect/language sets*, International Journal of Speech Technology, vol.18, no.3, pp.277–286, 2015.

[10] J. H. L. Hansen et al., *Dialect analysis and modeling for automatic classification*, Proceedings of 8th International Conference on Spoken Language Processing, 2004.