



Récapitulatif CSC 4538

Julien Romero

Python

À savoir

- Écrire un programme en Python
 - Déclaration de variables
 - Les types de base
 - Manipulation des chaînes de caractères
 - Listes, sets, dictionnaires
 - Structure de contrôle (if/elif/else, for, while)
 - Définition de fonctions
 - Les bases de la programmation objet en Python
 - Manipulation de fichiers

Scrapping

À savoir

- Fonctionnement HTTP
 - Structure de base d'une URL (chemin et requête)
 - Structure d'une requête (en particulier, le header)
- Fonctionnement d'une page web
 - Identifier la requête contenant d'information pertinente
 - Structure d'un document HTML
- Parcours d'un site web
 - Crawling
 - robots.txt
 - sitemap.xml

Visualisation de données

À savoir

- Connaître les types de diagrammes utiles et ceux à éviter
 - Texte, tableau, heatmap, nuage de points, lignes, diagramme à bande
- Les principes de Gestalt
 - Proximité, similarité, région commune, clôture, continuité, connectivité
- Attirer l'attention
 - Orientation, forme, longueur, épaisseur, taille, courbature, marquage, entourer, couleur, intensité, position spacial, mouvement

Introduction au traitement du langage naturel

À savoir

- Écriture d'une expression régulière
- Définitions classiques
 - Tokenisation
 - Vocabulaire
 - Normalisation : stemming vs lemmatisation, stop words
 - Part-of-speech
 - Dependency parsing
 - Named Entity recognition
- Vectorisation
 - Bag-of-words
 - TF-IDF

Recherche d'informations

À savoir

- Matrice d'incidence
- Construction d'un index inversé
- Algorithme pour faire des AND/OR sur un index inversé
- Index positionnel
 - Adaptation de l'algorithme pour les opérations booléennes
- Recherche ordonné
 - Jaccard
 - TF-IDF
 - Similarité cosinus

Systemes de recommandation

À savoir

- Matrice d'utilité
- Recommandation basée sur le contenu
 - Avantages et inconvénients
- Le problème du “cold-start”
- Filtrage collaboratif
 - Basé sur les utilisateurs/objets
 - Mesures de similarité (Jaccard, cosinus, corrélation de Pearson)
 - Prédiction du vecteur de caractéristique
 - Avantages et inconvénients
- Métriques d'évaluation
 - RMSE, Métrique sur le classement (Precision@K, Recall@K, MRR, MAP@K, NDCG@K)
- Filtrage collaboratif basé sur un model
 - Factorisation de matrice
 - SVD
 - Prédiction

Introduction au web sémantique

À savoir

- Current Web vs Semantic Web
- Knowledge Graph
- Semantic Web Stack
 - RDF
 - RDFS
 - OWL/Ontologies
 - Query: SPARQL
 - Rules: Reasoning
- Linked Data

Description de ressources en RDF/RDFS

À savoir

- RDF Triple/ RDF Graph
- Reification
- Inheritance:
 - `rdf:type` (class-instance relation)
- Complex types
 - `rdf:Bag`
 - `rdf:Sequence`
 - `rdf:Alt`
 - `rdf:List`
- N-aire Relations
- `rdfs:subClassOf` (class-superclass relation)
- `rdf:subPropertyOf` (property-superproperty relation)
- `rdfs:domain`
- `rdfs:range`

SPARQL

À savoir

- Read and write SPARQL queries
 - Optional
 - Filter

Introduction aux ontologies

À savoir

- OWL properties
 - Inverse, transitive, symmetric, functional, inverse functional
- OWL restrictions
 - Existential, Universal, hasValue, Cardinality
- Reasoning
 - Forward and Backward chaining
 - Writing rules with Jena