



IFT6285 (TALN)
Plan

Contact :
Philippe Langlais +1 514 343 61 11 ext: 47494
RALI/DIRO felipe@iro.umontreal.ca
Université de Montréal <http://www.iro.umontreal.ca/~felipe/>

Modus operandi

- Les ressources seront déposées sur Studium / *resources will be available on Studium* : <https://studium.umontreal.ca/course/view.php?id=176754>
- Les remises des travaux sont à faire sur Studium (des consignes seront données à cet effet). Les devoirs sont à remettre en groupe de deux ou seuls, les projets seront remis en groupes de deux ou trois personnes. / *All assignments will have to be submitted on Studium (more about this later). Devoirs must be done individually or in groups of two. Projects should be done in groups of 2 to 3 personnes.*
- Un espace de travail slack a été créé afin que vous puissiez communiquer avec moi. / *A slack workspace has been created, which is the best way to reach me* : <https://ift6285-automne2021.slack.com/>
- Les cours seront donnés en présence, selon les règles indiquées [ici](#). Je ferai de mon mieux pour enregistrer mon cours que je rendrai disponible depuis Studium. À priori, le cours sera donné en français, mais tous les énoncés de sujets (devoirs, projets et examen) seront également disponibles en anglais. / *Classes will be taught in person, according to the rules indicated [here](#). I will do my best to record my course and make it available from Studium. I will likely teach in French, but all assignments (devoirs, projects and exams) will also be available in English.*
- Une connaissance de base de Python est attendue. / *A familiarity with Python is expected.*

Executive summary

Teaching in **French** (might try in English). **Assignments** : 10 (devoirs) +2 (projects). **Exam** : one. **Language** : Python (at least for some librairies).

Matière / *Agenda*

Plan approximatif de notre session : d'autres sujets peuvent s'ajouter. / *Approximative plan of this semester, other subjects might come in.*

1. Introduction
2. Mains dans le cambouis / *Dirty hands*
 - quels langages de programmation / *which programming language?*
 - compter des mots / *counting words*
 - **devoir 1** : counting words in a large corpus
3. Modèles de langue / *Language models*
 - modèles discrets / *count-based models*
 - **devoir 2** : training and testing a model with kenlm
 - modèles neuronaux / *neural models*
 - **devoir 3** : to be determined
4. Approches distributionnelles / *Distributional Approaches*
 - vector space model, plongement de mots / *word embeddings*
 - **devoir 4** : Word embeddings with gensim (for instance)
 - **projet 1** : Kaggle like task
5. Algorithmique du texte / *Text Algorithmic*
 - comparaison de chaînes / *comparing strings* : distance d'édition / *edit-distance*, soundex, etc.
 - **devoir 5** : to be determined
 - recherche rapide / *fast search* : tables de suffixes / *suffix arrays*, FM-index
6. Étiquetage séquentiel / *Sequential Labelling*
 - étiquetage morpho-syntaxique / *Part-of-speech (POS) tagging*, chunking
 - **devoir 6** : to be determined
 - reconnaissance d'entités nommées / *named-entity recognition (NER)*
7. Analyse syntaxique / *Parsing*
 - théorie des langages / *language theory*
 - analyse en constituant et en dépendance
 - **devoir 7** : parsing with Spacy
8. Extraction d'information / *Information Extraction*
 - extraction de patrons

- extraction ouverte de relations
 - **devoir 8** : to be determined
 - **projet 2** : to be determined
9. Traduction automatique / *Automatique translation*
- approche statistique / *phrase-based models*
 - approche neuronale / *neural machine translation*
 - évaluation
 - **devoir 9** : to be determined
10. Question-Reponse / *Question Answering*
- **devoir 10** : to be determined
11. Apprentissage Analogique / *Analogical Learning*

Évaluation

- **10 devoirs** / *10 week assignments* : 3 x 10
 - remise au plus tard le vendredi de la semaine suivant l'énoncé / *assignment to be submitted by Friday next week*
 - des consignes claires accompagneront chaque devoir. Le respect de ces consignes conditionnera la note. / *Specific requirements will be asked, which will condition the score*
- **2 projets** / *2 projects* : 25 x 2
 - remise sur Studium en groupe de 2 à 3 personnes / *to be submitted on Studium in group of 2 to 3 persons*
 - un mois environ / *a month or so*
 - Vos projets seront notés d'après un rapport (pdf) de quelques pages. La note dépendra notamment : / *Projects will be scored thanks the companion report. We will search for :*
 - de votre curiosité / *curiosity*
 - de la clarté de votre rapport / *clarity*
 - du contenu informationnel / *substance*
- **examen** final / *final exam* : 20
 - question de compréhension sur des choses vues en cours / *Compréhension questions on subjects discussed in course*