

IFT714 : Traitement automatique des langues naturelles

Description du projet

Le projet est l'occasion d'étudier un problème NLP intéressant **de votre choix** ou de la [liste de sujets / problèmes proposés](#). Le projet doit être une **étude de recherche expérimentale** généralement dans le contexte d'un ensemble de données textuelles du monde réel.

Table des matières

1. Projet de recherche expérimentale
2. Livrables (Langue: Français ou Anglais)
 - 2.1. Proposition de projet (date limite : 19 février 2024, 1 page)

Tout membre du groupe peut soumettre la proposition de projet sur Moodle. Notez qu'il s'agit d'un dossier de soumission de groupe et non d'un dossier de soumission individuel.
 - 2.2. Rapport d'avancement ou de progression (~~18~~ 22 mars 2024)
 - 2.3. Présentation (10 avril et 15 avril 2024 en classe)
 - 2.4. Rapport final (vendredi 19 avril 2024)
3. Critères d'évaluation
4. Format/Style
5. Logistique
6. Prochaines étapes
7. Liens utiles vers des ensembles de données et des tâches (de Chris Manning)
8. Collecte de vos propres données
9. Recommandations sur la rédaction de rapports
10. Exemple de projet de cours
11. Exemples similaires aux documents de recherche expérimentale

1. Projet de recherche expérimentale

Pour le projet de recherche expérimentale, l'enquête peut commencer par la lecture de 10 à 15 articles connexes de conférences / revues sur NLP (ACL, NAACL, EMNLP, EACL, AAAI, ICWSM,

COLING, CoNLL, Transactions of the ACL, Computational Linguistics, etc.). Commencez à lire des articles liés à votre sujet préféré ou à un [sujet](#) proposé ([consultez les ressources de la Semaine 2 dans Moodle pour savoir comment lire un article](#)). Une fois qu'une itération initiale de la lecture de la littérature et de la discussion au sein des membres de l'équipe est terminée, vous devez filtrer ou sélectionner les articles pertinents liés à vos objectifs. Mais quels sont vos objectifs ? Eh bien, la revue de la littérature devrait vous permettre d'identifier clairement vos objectifs : quelles sont les entrées et les sorties de la tâche ? Qu'essayez-vous de réaliser ? Quel(s) dataset(s) (jeu(x) de données) allez-vous utiliser pour entraîner votre modèle ? Auriez-vous besoin d'un prétraitement ? Comment évaluez-vous votre succès, votre échec (métrique d'évaluation) ? À quoi ressemble le succès ? Si vous faites une analyse, quelle est votre hypothèse ? Comment vos expériences confirmeront-elles ou infirmeront-elles votre hypothèse ? **Les réponses à ces questions vous aideront à rédiger le rapport de proposition.**

Une recherche expérimentale peut impliquer l'application d'un modèle d'apprentissage existant à une nouvelle tâche proposée (qui n'a jamais été abordée auparavant) ou à une tâche où le modèle existant n'a jamais été essayé, mais où vous pensez qu'il pourrait potentiellement fonctionner pour des raisons spécifiques; ou la mise en œuvre d'un nouveau modèle; ou présenter une nouvelle variante pour un modèle existant; ou proposer un nouveau schéma d'optimisation ou d'évaluation; ou fournir une analyse expérimentale de différents modèles sur un nouvel ensemble de données. Des figures et des tableaux résumant les architectures conçues et affichant et analysant les résultats devraient être inclus dans les rapports.

Les projets expérimentaux devraient comprendre :

- un examen approfondi de la ***littérature***, donnant un aperçu du sujet général;
 - Lire des articles qui sont liés, c'est-à-dire des articles qui
 - s'attaquent au même problème mais en utilisant des approches différentes;
 - OU utilisent des techniques similaires pour résoudre différents problèmes (si vous êtes intéressé par une technique spécifique);

- OU comme l'un est un travail de suivi de l'autre;
- OU ont des points de vue opposés sur certains problèmes; etc.
- Vous pouvez commencer par rechercher des articles liés à votre sujet dans les récentes conférences de l'[Association of Computational Linguistics \(ACL\)](#), comme [ACL23](#), [ACL22](#), [EACL21](#), [EMNLP21](#), [AAACL20](#), [ACL21](#), [SemEval23](#), [SemEval22](#), [SemEval21](#) et [SemEval20](#) et plus généralement dans [l'ACL Anthology](#) (un référentiel de nombreux actes de conférences NLP indexés) ou [Google Scholar](#) ou [arXiv.org](#), ou <https://dl.acm.org>. Trouvez les articles pertinents et vous pouvez explorer les travaux connexes qu'ils citent (dans la section Références), ainsi que les articles les citant en utilisant par exemple [Google Scholar](#) (lien « cité par » sous le titre et l'extrait de l'article). **La revue de la littérature fera partie des quatre produits livrables liés au projet (proposition, progrès, présentation et rapports finaux), même si la plupart des efforts concernant cet examen de la littérature seront effectués pendant que vous travaillerez sur les rapports de proposition et de progrès.**
- **L'expérimentation empirique** de modèles/architectures sur une tâche ou un problème particulier.

2. Livrables (Langue: Français ou Anglais)

Le projet impliquera les tâches de:

- décider d'un sujet spécifique;
- lire la littérature pertinente;
- mettre en œuvre certaines des idées
- rédaction d'un article original de **4** pages suivant les modèles officiels de style ACL (détails sur le format ACL dans la section Format / Style ci-dessous);
 - passer en revue de façon critique la littérature;
 - présenter votre approche, vos résultats, vos analyses, vos constats.

2.1. Proposition de projet (date limite : 19 février 2024, 1 page)

La proposition de projet ne doit pas dépasser une page en [format *ACL](#) (pas besoin d'inclure le résumé (abstract)).

Tout membre du groupe peut soumettre la proposition de projet sur Moodle. Notez qu'il s'agit d'un dossier de soumission de groupe et non d'un dossier de soumission individuel.

Soumettez uniquement le fichier « .pdf ».

1. Un énoncé clair du problème abordé, c'est-à-dire, Quel est le problème?
 - Exemples
 - Définition (entrées-sorties)
 - Une figure qui montre explicitement l'entrée et la sortie de la tâche ou du système d'apprentissage pourrait également aider
 - À quoi ressemblerait la solution?
2. Pourquoi le problème est-il important? (c'est-à-dire, pourquoi devrais-je m'en soucier?)
 - Si cela était résolu, qu'est-ce que cela rendrait possible?
 - Si cela était résolu, le monde serait-il un endroit plus heureux?
3. Pourquoi cette tâche n'est-elle pas triviale? c'est-à-dire des défis!!
 - Ici, il peut être utile de présenter un exemple bien choisi pour donner un aperçu des complexités - un exemple est un exemple facile à énoncer, mais pas évident comment résoudre. (Par exemple, montrer que les solutions évidentes ne fonctionnent pas)
4. Les personnes (chercheurs, développeurs, techniciens) qui ont vraiment besoin de résoudre ce problème: Que font-ils? Et quelle sont les limites? En quoi leur approche est-elle différente/similaire à ce que vous allez proposer?

- Cela devrait être un **bref** résumé critique des **TYPES** de techniques, de leurs mérites et de leurs limites, **pas** une réplique ennuyeuse de tout ce que vous avez lu.

5. Que proposez-vous pour résoudre le problème?
6. Inclure des références (travaux connexes) et les citer de manière appropriée. La section Références doit apparaître **après le contenu de la proposition de 1 page**.
7. Énumérer les membres de l'équipe en tant qu'auteurs (sous le titre du projet)
8. Indiquez la contribution de chaque membre de l'équipe à la production de la proposition ainsi que leurs futures tâches prévues dans le projet. **Cela devrait apparaître après la page Référence.**
9. Format - Format ACL ([template](#), [overleaf](#))
10. Après la page Référence, indiquez le dataset que vous prévoyez d'utiliser et assurez-vous qu'il est disponible et que vous y avez accès. Indiquez également sa taille.
11. Après la page Référence, indiquez les ressources informatiques potentielles auxquelles vous avez accès pour exécuter vos expériences/méthodes.

2.2. Rapport d'avancement ou de progression (~~18~~ 22 mars 2024)

Limite de 2 pages (excluant les références et les contributions)

Ce rapport devrait vous assurer que vous progressez dans votre projet.

Le rapport d'avancement sera rédigé dans le même format *ACL que celui utilisé pour rédiger la proposition de projet. Il devrait contenir une version préliminaire des sections de votre rapport final, comme la revue de littérature ou les travaux connexes (related work). Cela vous facilitera la rédaction de votre rapport final. Vous constaterez peut-être que vous pouvez réutiliser des parties de votre proposition de projet dans votre rapport d'avancement. C'est correct tant que vous vous assurez d'intégrer les commentaires que vous avez reçus sur votre proposition et d'élargir la section des travaux connexes de manière appropriée avec les nouvelles lectures pertinentes que vous avez faites depuis.

Soyez concis : allouez moins d'espace à la réintroduction du problème (les entrées-sorties doivent être claires, vérifier les commentaires) et plus d'espace à ce que vous avez accompli, c'est-à-dire votre architecture de méthode (c'est-à-dire la conception détaillée du modèle, pas seulement en spécifiant que vous allez utiliser de l'apprentissage automatique ou en affichant un organigramme d'exploration de données) et vos expériences et évaluations préliminaires.

Veuillez inclure les informations suivantes dans votre rapport d'étape :

- **Titre** : Préfixé par le numéro de groupe : p. ex. Groupe 19 : SemEval Tâche 7 - Détection des fausses nouvelles
- Uniquement les **noms des membres du groupe** avec leurs identifiants. Les e-mails et les informations sur le département ne sont pas nécessaires (cela vous fera gagner de l'espace)
- **Nom du fichier PDF** : Veuillez utiliser ce format Group_XX-Progress_Report.pdf
- Indiquez **la contribution de chaque membre de l'équipe à l'avancement jusqu'à présent** ainsi que **leurs futures tâches** prévues dans le projet. **Cela devrait apparaître après la page Référence**

Les composantes du rapport d'étape :

- **Énoncé clair et concis du problème** (se référer aux exigences de la proposition et aux commentaires à ce sujet, inclure un bref exemple d'entrée et de sortie de la tâche)

- **Travail connexe: une enquête plus** approfondie des travaux connexes antérieurs basée sur vos lectures récentes et les commentaires, une meilleure formulation des **types de techniques, leurs avantages, leurs limites**, ce qu'elles modélisent, et **dans quelle mesure vous abordez le problème de manière similaire ou différente et pourquoi**.
- **Approche ou méthodologie:** Vous devez détailler votre approche pour résoudre le problème. C'est ici que vous décririez la conception ou l'architecture, les méthodes ou les algorithmes de votre modèle (pas besoin de prendre une demi-page pour décrire un algorithme, soyez concis).
 - Soyez précis lorsque vous décrivez vos principales approches. Vous pouvez inclure des équations clés (seulement si c'est vraiment pertinent), des figures (votre propre figure pour votre architecture conçue, pas celle de quelqu'un d'autre). Vous pouvez, si vous voulez, reporter la tâche de création de figures explicatives, qui pourrait vous consommer du temps et de l'espace, à votre rapport final, mais assurez-vous de décrire votre approche de la manière la plus claire possible.
 - Décrivez **une ligne de base ou de référence (BASELINE)** à laquelle vous comparez. Une base de référence peut être une méthode issue du travail connexe sur ce problème particulier ou une ou plusieurs méthodes simples que vous exécutez et avec lesquelles vous comparez votre approche (par exemple, un simple NN, Naive Bayes, SVM, etc.). Vous n'avez pas à décrire le fonctionnement des algorithmes de baselines connus comme le NN, Naive Bayes, SVM, KNN, etc.
 - Si une partie de votre approche est originale, dites-le clairement. Pour les modèles et les techniques qui ne sont pas les vôtres, fournissez des références.
 - Si vous utilisez un code que vous n'avez pas écrit vous-même, indiquez-le clairement et fournissez une référence ou un lien. Lorsque vous décrivez quelque chose que vous avez codé vous-même, dites-le clairement.
- **Expériences**
 - **Données** : Décrivez le(s) jeu(s) de données que vous utilisez ainsi que les références. Assurez-vous d'expliquer clairement comment vous utilisez les exemples de l'ensemble de données au cas où l'entrée et la sortie décrites dans

l'énoncé du problème ne correspondent pas exactement à l'entrée et à la sortie de l'ensemble de données que vous utilisez, ou si l'ensemble de données est généralement utilisé pour quelque chose de différent ou de plus général (par exemple, jeu de données de tweets généraux) que ce pour quoi vous l'utilisez (par exemple, Tâche liée au Covid-19).

- **Méthode d'évaluation** : Décrivez les paramètres d'évaluation que vous utiliserez ou avez utilisés, ainsi que tout autre détail nécessaire pour comprendre votre évaluation.
- **Détails expérimentaux** : Veuillez expliquer comment vous avez mené certaines de vos expériences (par exemple, configurations de modèles, taux d'apprentissage, temps de formation, etc.). Cela devrait fournir de manière concise des détails techniques, pas un récit.
- **Résultats** : Indiquez les résultats quantitatifs que vous avez obtenus jusqu'à présent. Utilisez un tableau ou un graphique pour comparer plusieurs résultats et comparer avec vos baselines. Commentez vos résultats quantitatifs. Sont-ils ce à quoi vous vous attendiez? Mieux que ce à quoi vous vous attendiez? Pire que ce à quoi vous vous attendiez? Pourquoi pensez-vous que c'est le cas? Qu'est-ce que cela vous dit sur ce que vous devriez faire ensuite?
- **Travaux futurs** : Décrivez ce que vous prévoyez de faire pour le reste du projet et pourquoi (autres expériences, raison et évaluation potentielle).
- **Références** : Utilisez BibTeX pour produire vos références.

Vous serez jugé sur les différentes parties décrites ci-dessus.

Ainsi, vous serez évalué sur vos **progrès**. L'équipe a-t-elle bien progressé dans le projet? Vous devriez avoir fait environ la moitié du travail de votre projet. Vos progrès doivent montrer que vous avez configuré vos données, votre modèle de référence et votre métrique d'évaluation, que vous avez effectué des expériences pour obtenir des résultats préliminaires. En dehors de cela, les « bons progrès » dépendent de divers facteurs (par exemple, si votre modèle est implémenté à partir de zéro (vraiment rare) ou basé sur une base de code existante).

Vous serez également évalué sur la compréhension: les progrès montrent-ils une bonne compréhension du problème, des tâches, des méthodes, des mesures et du contexte de recherche?

Un autre élément de l'évaluation est la qualité de la rédaction : le rapport communique-t-il clairement ce que vous avez fait et pourquoi, en fournissant les renseignements demandés, à un niveau de détail approprié (compte tenu de la limite de pages)?

En résumé, un rapport d'étape devrait contenir :

- **Un énoncé bref et clair du problème**
- **Une étude plus approfondie des travaux connexes** antérieurs, des types de techniques, de **leurs avantages**, de **leurs limites**, de ce qu'elles modélisent, **dans quelle mesure** abordez-vous le problème différemment ?
- Plus de détails factuels sur la méthodologie suivie pour aborder le problème.
- Décrivez l'architecture du système (p. ex. figure), les expériences en cours et celles qui sont prévues.
- Indiquez les critères d'évaluation que vous utilisez ou utiliserez.
- Affichez quelques résultats préliminaires initiaux.

2.3. Présentation (10 avril et 15 avril 2024 en classe)

- Présentez les points **abordés dans les rapports précédents (proposition et état d'avancement) et les progrès réalisés depuis le dernier dérivable** (85% à 90% du travail de mise en œuvre, résultats, analyse

2.4. Rapport final (vendredi 19 avril 2024)

Limite de pages : **4 pages** excluant les références, l'annexe et les contributions.

Exigences (le rapport final comprendra une description complète des travaux entrepris dans le cadre du projet)

- Résumé : **Un très bref** résumé du problème papier, de la méthode utilisée pour le résoudre et des principaux résultats.
- Un énoncé clair du problème traité
 - Il est souvent utile d’inclure une figure qui montre explicitement l’entrée et la sortie du système
- Motivation - c’est-à-dire, pourquoi ce problème est-il intéressant et difficile?
- Matériel de base nécessaire, examen des travaux antérieurs et de leurs limites, et positionnez votre travail dans ce domaine (en quoi est-il différent? aborde-t-il une limitation).
- Un énoncé clair des solutions techniques utilisées pour résoudre le problème, de leur succès et des raisons de leur succès.
 - Utilisez des termes clairement définis. Il est bon d’utiliser des intuitions pour *motiver* une idée, mais ensuite utilisez des énoncés précis pour énoncer vos revendications, puis étayez ces affirmations avec des résultats empiriques significatifs.
- Une analyse approfondie des résultats et des points à retenir
- Identification des problèmes restants et des orientations futures de la recherche.
- en annexe (après les références) : indiquez les bibliothèques et les logiciels utilisés et les contributions de chaque membre tout au long du projet.
- **Autres livrables :**
 - **Le code** doit être soumis avec un fichier Lisez-moi expliquant comment l’exécuter.
 - Incluez tout le code de projet écrit ou adapté par vous.
 - N’incluez pas le code source complet des paquets prêts à l’emploi que vous avez utilisés sans les adapter (par exemple, CoreNLP ou PyTorch).
 - Incluez **l’ensemble de données** si possible, sinon fournissez un lien vers celui-ci.

- Le code doit être aussi propre et commenté que possible. Les marques peuvent être déduites en cas de manque de clarté du code.

En bref, il devrait répondre aux questions:

Quel a été le problème sur lequel vous avez travaillé, pourquoi est-ce important, qu'avez-vous appris, conçu, analysé, comparé et pourquoi ces résultats sont-ils importants, quelles conclusions pouvons-nous tirer?

3. Critères d'évaluation

Notez qu'étant donné que vous n'avez que quelques semaines pour travailler sur votre projet, il **n'est pas** nécessaire que votre méthode batte les performances de pointe (state-of-the-art) pour la tâche choisie, ou fonctionne mieux que les méthodes précédentes. Mais il devrait au moins montrer les performances largement attendues des types de méthodes que vous utilisez compte tenu de la taille du jeu de données et des ressources disponibles. Dans tous les cas, votre article devrait essayer de fournir un raisonnement expliquant le comportement de votre modèle. Vous devrez fournir une analyse qualitative, qui sera utile pour soutenir ou tester vos explications. Cela sera particulièrement important lorsque votre méthode ne fonctionne pas aussi bien que prévu.

Vos rapports de projet seront évalués en fonction de

- *Quantité de travail* consacrée au projet
- *Qualité de rédaction et Clarté.* Votre document démontre-t-il une bonne compréhension et donne-t-il une bonne analyse du défi sous-jacent? Est-il bien écrit et bien organisé? Utilise-t-il à bon escient des exemples pour illustrer le problème et les solutions?
- *Originalité.* Votre document fait-il preuve d'une réflexion et d'une considération indépendantes, identifie-t-il les problèmes non résolus et propose-t-il des solutions initiales, ou propose-t-il de nouvelles approches aux problèmes existants?
- *Solidité et substance* (pour la pertinence des techniques que vous avez utilisées dans le cadre de votre problème)
- *Évaluation* (rigueur de votre évaluation, comparaison significative)

4. Format/Style

- Tous les rapports DOIVENT UTILISER les modèles officiels de style ACL, qui sont disponibles sous forme de [modèle Overleaf](#) et également [téléchargeables directement](#) (Latex et Word).
- Toutes les limites de pages sur les rapports excluent les références (vous pouvez avoir autant de références que vous le souhaitez, la limite est sur le contenu principal)
- INCLUEZ vos noms sur le rapport (par défaut, les modèles peuvent être anonymes). Les noms et identifiants des membres du groupe sont suffisants (pas besoin de préciser le département. Cela vous fera gagner de l'espace.)

5. Logistique

- Chaque projet sera réalisé par une équipe de **3 étudiants**. Pour chaque rapport, vous devez inclure une annexe à la fin du document, qui indiquera la contribution de chaque membre du groupe (par exemple, lire et synthétiser 3 à 5 articles, discuter des travaux et de la méthodologie connexes avec les membres du groupe, concevoir une architecture ou une nouvelle perspective pour comparer les articles, implémenter et exécuter le code, analyser les résultats, etc.). La présentation finale devrait contenir une diapositive sur la contribution de chaque membre au projet (jusqu'à ce point et peut-être potentiellement par la suite). Tous les membres du groupe doivent être présents lors de leur présentation de groupe. **L'instructeur peut poser des questions aux différents membres de l'équipe après la présentation concernant leurs contributions respectives.**
- **Utilisation de ressources externes :**
 - Vous pouvez utiliser n'importe quel framework de deep learning que vous préférez (PyTorch, TensorFlow, etc.)

- Plus généralement, vous pouvez utiliser n'importe quel code, bibliothèque, etc. existant et consulter des articles, des livres, des références en ligne, etc. pour votre projet. Cependant, vous devez citer vos sources dans votre article et indiquer clairement quelles parties du projet sont votre contribution et quelles parties ont été mises en œuvre par d'autres.
- En aucun cas, vous ne pouvez consulter le code d'un autre groupe ou incorporer son code dans votre projet (en supposant que vous choisissiez le même problème ou la même tâche).

6. Prochaines étapes

- Inscription de groupe **26 janvier**
 - Équipes avec 3 noms de membres du groupe
- Peut inclure **le sujet du projet** si vous avez une idée initiale ou des doutes.
- Proposition de projet (**19 février**)
- Rapport d'avancement/progression (**17 mars**)
- Présentation (**12 avril et 14 avril**)
- Rapport final (**19 avril**)

7. Liens utiles vers des ensembles de données et des tâches (de Chris Manning)

- Une collection de documents avec du code sur de nombreuses tâches NLP.
<https://paperswithcode.com/sota>
- Un référentiel pour suivre les progrès dans NLP, y compris les listes des ensembles de données et les performances actuelles de pointe pour les tâches NLP les plus courantes.
<https://nlpprogress.com/>
- Microsoft Research dispose d'une collection de jeux de données (regardez sous l'onglet « Dataset directory », recherchez l'informatique ou les sciences sociales et effectuez une

recherche avec les mots-clés « text » et « acl ») :

<http://research.microsoft.com/en-US/projects/data-science-initiative/datasets.aspx>

<https://msrpendata.com/datasets?domain=COMPUTER%20SCIENCE>

<https://msrpendata.com/datasets?domain=SOCIAL%20SCIENCE>

- Une petite liste de jeux de données standard bien connus pour les tâches NLP courantes : <https://machinelearningmastery.com/datasets-natural-language-processing/>
- Liste alphabétique des ensembles de données textuelles libres ou du domaine public : <https://github.com/niderhoff/nlp-datasets>
- Wikipédia dispose d'une liste d'ensembles de données textuelles d'apprentissage automatique, tabulée avec des informations utiles telles que la taille des ensembles de données.
https://en.wikipedia.org/wiki/List_of_datasets_for_machine-learning_research#Text_data
- Kaggle a de nombreux ensembles de données, bien que certains d'entre eux soient trop petits si vous utilisez Deep Learning. Essayez de rechercher 'nlp'.
<https://www.kaggle.com/datasets>
- Datahub a beaucoup de jeux de données, bien que tous ne soient pas axés sur le NLP ou l'apprentissage automatique.
<https://datahub.io/collections>
- Un script pour rechercher un mot-clé dans les articles arXiv et extraire des informations importantes telles que les mesures de performance d'une tâche.
<https://huyenchip.com/2018/10/04/sotawhat.html>

8. Collecte de vos propres données

La collecte de données est souvent un processus long et compliqué qui est plus difficile qu'il n'y paraît. Compte tenu du délai limité, ce n'est pas recommandé généralement de collecter vos propres données. Si vous souhaitez vraiment collecter vos propres données, assurez-vous de planifier correctement le temps de collecte des données dans votre projet. N'oubliez pas que

vosre projet peut avoir une composante de modèle d'apprentissage substantielle, donc si vous passez tout votre temps sur la collecte de données et aucun sur la construction et l'évaluation de votre modèle, ça ne sera pas possible d'avoir une bonne note.

9. Recommandations sur la rédaction de rapports

- Essayez de rendre votre rapport FACILE à lire.
- Pour le rapport final, assurez-vous d'inclure un aperçu au début (fin de l'introduction), qui décrit ce que le document décrira, section par section.
- Incluez des exemples simples (ou mieux, un seul exemple simple tout au long), pour aider à illustrer les idées.
- Une image vaut (au moins) mille mots. Utilisez des figures, des organigrammes, des graphiques, le cas échéant.
- Le contenu doit être structuré et fluide. Cela ne devrait PAS être un core-dump de tout ce que vous avez lu lorsque vous regardiez des choses liées à X.
- Si vous donnez une description de haut niveau d'un algorithme, assurez-vous d'indiquer explicitement son entrée et sa sortie.
- De nombreux algorithmes ont un flux d'informations, d'un sous-programme à l'autre. Fournissez une ou plusieurs figures pour clarifier les idées.
- Aussi, relisez votre rapport. Evitez de remettre un rapport qui contient beaucoup de fautes de frappe faciles à corriger, des chiffres illisibles, des citations manquantes, etc.
- Si vous décrivez un algorithme précis, vous devez donner les formules réelles, en utilisant des termes bien définis, dans le rapport.
- Votre rapport doit être "self-contained". Vous êtes autorisé à copier des figures provenant d'autres sources (s'ils sont correctement crédités). Mais si vous le faites, assurez-vous de définir les termes qui apparaissent dans cette figure !

- Et utilisez les numéros de section et les numéros de page !

10. Exemple de projet de cours

- **Sequence-to-Sequence Generative Argumentative Dialogue Systems with Self-Attention**
<https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1194/reports/custom/15844523.pdf>

11. Exemples similaires aux documents de recherche expérimentale

- ANA at SemEval-2019 Task 3: Contextual Emotion detection in Conversations through hierarchical LSTMs and BERT
<https://arxiv.org/pdf/1904.00132.pdf>
- Automatic Dialogue Generation with Expressed Emotions
<https://www.aclweb.org/anthology/N18-2008.pdf>
- EDA: Easy Data Augmentation Techniques for Boosting Performance on Text Classification Tasks
<https://arxiv.org/pdf/1901.11196.pdf>
- ANA* at SemEval-2020Task4: mUlti-task learNing for cOMmonsense reasoNing (UNION)
<https://www.aclweb.org/anthology/2020.semeval-1.45.pdf>
- The Surprising Performance of Simple Baselines for Misinformation Detection
 € <https://arxiv.org/pdf/2104.06952.pdf>

Crédit à Chris Manning et Russ Greiner, la majeure partie de la description du projet a été adaptée et utilisée à partir de leur NLP CS224n 2020 avec DL et 2008 Cmpu651 PGM course, descriptions de projet, respectivement; Le mérite revient également à Diyi Wang et David Bamman pour leurs matériaux.