

Pauta Hito 1

El objetivo del hito 1, tanto para el regular track como para el research track, es realizar una investigación preliminar sobre la task a trabajar a lo largo del semestre.

Los puntos que deben desarrollar para el hito 1 son:

1. **Motivación:** aquí deben motivar la task escogida. Ejemplo: Machine translation es importante porque permite comunicar a personas que no hablan la misma lengua.
2. **Definición formal de la task:** aquí deben definir formalmente en qué consiste la task escogida. ¿Cuales son los inputs y que formato tienen? ¿Cuales son los outputs y que formato tienen? Ejemplo: para machine translation el input es texto en el idioma fuente y el output es texto en el idioma destino.
3. **Datasets:** aquí deben describir los datasets o corpora que usarán en el proyecto. Es muy importante que obtengan los datos de training, validation y testing en esta etapa del proyecto. Deben mostrar que ya tienen los datos. Por favor eviten proyectos donde vean problemas para conseguir los datos.
4. **Métricas de evaluación:** aquí deben explicar las métricas de evaluación de la task. Ejemplo: en machine translation se usa BLEU que es una métrica en el rango [0-1] y que representa la precisión de palabras de la traducción generada respecto a la traducción de referencia. Se espera que entiendan bien la métrica usada. Deben describirla matemáticamente en su informe.
5. **Modelo a implementar y baseline de comparación:** aquí deben describir el modelo escogido para desarrollar durante el proyecto. Deben citar al paper de referencia que describe el modelo. También deben proponer un baseline, que es un modelo simple, contra el cual van a contrastar el modelo escogido. Por ejemplo: si la tarea es sentiment analysis, el modelo escogido puede ser una arquitectura compleja que usa LSTMs, CNNs, o Transformers, mientras el baseline es un modelo bag-of-words con una regresión logística. Como cuerpo docente entendemos que esto puede ser difícil de definir a esta altura. Lo importante es que sean capaces de proponer un plan tentativo.

Los entregables del hito 1 son:

Informe

Un informe breve con una sección por cada uno de los puntos de más arriba más un abstract que resuma todo el informe. El informe debe estar escrito en LaTeX usando el template de ACL. Les recomiendo usar OverLeaf con el siguiente template:

<https://www.overleaf.com/latex/templates/instructions-for-acl-2019-proceedings/wvjpcrzdsqz>
[v](#)

Para las referencias deben ponerlas en el archivo acl2019.bib del template de overleaf en formato BibTex. La forma más fácil es buscar el paper en Google Scholar y hacer click en las “cremallas” para poder bajar la referencia en formato BibTex. Pregunten en Ucourses si tienen dudas.

Si algún grupo tiene una razón de peso para no usar LaTeX que lo comunique al cuerpo docente.

Presentación

Una presentación de máximo 7 minutos de duración que aborde todos los puntos de más arriba. El tiempo lo vamos a medir estrictamente. Preocupense de no pasarse en el tiempo. El orden de las presentaciones va a ser aleatorio a excepción de los grupos que quieran presentar con anterioridad.

Los grupos del regular track presentarán en horario de clases, mientras que los grupos del research track presentarán en otro horario (lo más probable es que sea un Miércoles a las 14PM).

Entrega

Tanto el informe como la presentación deberán ser subidas a Ucourses en formato PDF (por favor conviertan sus PPTs a PDF). El plazo de entrega es el lunes 09/09/2019 a las 23:59.

Sobre la dificultad de las tasks

Existen tasks que son mucho más complejas que otras, ejemplo: text classification es más simple que question answering. Para los grupos que escojan tareas más simples se esperará que hagan mucho más experimentos durante el hito 2.