Este documento pretende recoger los contenidos de las reuniones con Rocío: preguntas, comentarios, anotaciones, tareas que hacer, cosas pendientes semana a semana, revisiones, y puntos a mejorar, entre otras cosas.

# Primera reunión (27/10/22)

## Objetivos de la reunión

* Presentación y primera toma de contacto con el tema del trabajo

## Preguntas para esta reunión

* ¿Estructura de la memoria? 🡪 ver tfgs antiguos
* Vías de comunicación 🡪 por correo
* ¿Llevar un registro de horas de cada cosa en lo que trabajo? 🡪 sí, pero no es necesario que sea muy exhaustivo

## Conclusiones de la reunión

Me enviará material de lectura, tfgs antiguos de referencia para que les eche un vistazo, un set de 100 moléculas con los que poder realizar el análisis, y código para probar en Python.

- Ir guardando la bibliografía que vea, y si es una página web apuntar el día de visita (ir guardando esto para luego no quedarme vendido)

- Páginas webs que usaré para las cadenas SMILES: sigmaaldrich.com / scifinder / pubchem (esto está pensado para investigación, los compuestos comerciales no los suele tener). Yo tengo que comparar scifinder y sigmaaldrich principalmente.

- [Rdkit va muy bien para cosas básicas, pero cuando se le pide un poco más, no funciona bien. Algunas moléculas de organometalica ni siqueira las acepta por ser químicamente inviables. Este paquete lo podemos usar para probar y hacer una introducción de que existe más de un paquete, pero vamos a trabajar con Openbabel. (no tengo que comparar los paquetes como tal, es simplemente para probarlo)

Si dibujamos la misma molécula con ambos paquetes, dan un dibujo distinto, que no tiene por qué estar mal.

Openbabel tiene un método ‘canonical’ que si le metes una cadena SMILEs, lo convierte a la secuencia canónica, con la idea que todos los sinónimos de la misma molécula, debería darte el mismo canónico. Esa es la idea, pero no funciona del todo bien.

Mi tarea es crear un método canonical que mejore aunque sea mínimamente eso, específicamente para moléculas organometalicas.

Cosas que hacen mal todos los paquetes: dibujar. Para las moléculas convencionales va muy bien, para las orgánicas/inorgánicas también; pero las organometalicas le cuesta. SciFinder por otro lado, hace los dibujos mucho mejor, pero es de pago y no tenemos acceso al código de como lo hacen.

Hay un paquete que usan los químicos para dibujar, que es chemdraw o algo así. Pero solo funciona con RDkit, y como rdkit no se traga las moléculas organometalicas, no podemos usar esto.]

## Tareas nuevas que hacer

* Leer papers y material 🡪Hecho
* Repaso de química básica 🡪 Hecho
* Estudiar y escoger metodología de trabajo (tradicional o agil) 🡪 **Por justificar**
* Recursos: que costo tiene mi proyecto 🡪**Por hacer**
* Analisis de riesgos, planes de contingencia 🡪**Por hacer**
* Ir guardando la bibliografía que vaya viendo 🡪 viene bien usar gestores como Zotero**.**
* Algo para contabilizar las horas que le echo al tfg🡪 Uso Clockify

## Próxima reunión

La próxima reunión está fechada para el 09/11/22

# Segunda reunión (16/02/23)

La reunión para el 09/11/22 quedó a la espera debido a las asignaturas del primer cuatrimestre y las vacaciones de Navidad.

## Objetivos de la reunión

* Repasar los objetivos del trabajo en líneas generales
* Establecer la metodología de desarrollo
* Resolver las dudas surgidas tras las lecturas iniciales
* Problemas instalando algunos paquetes de python para probar el código

## Preguntas para esta reunión

* ¿Hasta qué punto necesito aprender SMILES, es solo para describir compuestos, o también reacciones?
  + Respuesta: lo justo solo para describir compuestos y moléculas
* ¿Cómo busco artículos e información buena? ¿Cómo accedo a plataformas tipo ieeexplore o ACSPublications?
  + Respuesta: Normalmente la ugr tiene convenios para poder acceder a algunos papers. Pero lo más fácil es usar Google Scholar y buscar por ahí. Todo lo que vea/lea (aunque sea por curiosidad, o solamente el abstract, o directamente no lo use para la memoria) tengo que guardarlo.
* Dijiste que tenias una idea de como podría ser esa forma canónica. ¿Eso lo tengo que desarrollar yo, o los químicos con los que trabajas me lo proporcionan?
  + Respuesta: mirar conclusiones
* ¿Cómo accedo a SciFinder, necesito registrarme con una cuenta ugr?
  + Respuesta: en principio si, se supone que es gratis, pero tengo que usarlo desde la VPN de la ugr. 🡪 **No me funciona, necesito un formulario especial para hacer el registro.**

## Conclusiones de la reunión

- Hay varias representaciones lineales modernas. SMILES, InChI y SELFIES. SMILES es la más usada, pero con el problema de que una misma molécula se puede escribir de muchas formas distintas. InChI va muy bien para la máquina, pero no es nada legible por los humanos. Y SELFIES, tiene la ventaja de que una molécula se representa de una única manera, y todas las representaciones que genera son químicamente viables; cosa que SMILES no hace.

A la hora de dibujar las moléculas, específicamente en el campo de la organometalica, no se dibujan para nada bien. Si le meto un código SMILES determinado me lo dibuja de una manera; si lo modifico (reordenando la cadena) aun siendo el mismo compuesto, me lo dibuja de otra manera distinta. Como en cada página (base de datos) tienen un código diferente, no hay nada unificado.

- Por tanto, tengo que analizar esas diferencias entre la misma molécula dependiendo de donde la saque (webs).

- Determinar un sistema que me genere, a partir de cualquier sinónimo SMILES de la misma molécula, un único SMILES canonico. Internamente Openbabel monta un grafo. La idea es a partir del SMILES de entrada, montarlo en un grafo; y escribir ese grafo de manera canonica. Aquí la complejidad está en ver cómo recorro ese grafo. Esto me tendré que inventar algunas reglas. (p.ej, usando el nº atomico)

Cadena original 🡪 Grafo (esto ya lo hace openbabel) 🡪 Generar cadena canonica

## Tareas nuevas que hacer

* Mirar Google Colab e intentar montar ahí el entorno de pruebas si me sigue fallando el openbabel. 🡪 Hecho
* Hacer una tabla resumen con las diferencias entre los códigos SMILES y dibujos para una misma molécula en las distintas webs. Usar también Rdkit, hay algunas moléculas que no soporta. Esto para la introducción/motivación. 🡪 **Por completar**
* Escribir la introducción al trabajo entorno a lo anterior (química informática, sistemas de representación, organometalica…) 🡪 **Por completar**
* En general, ir rellenando las partes iniciales de la memoria.

## Tareas pendientes

* Estudiar y escoger metodología de trabajo (tradicional o agil) 🡪 **Por justificar**
* Recursos: que costo tiene mi proyecto 🡪 **Por hacer**
* Analisis de riesgos, planes de contingencia 🡪 **Por hacer**

## Tareas a futuro

* Una vez tenga la introducción medio que, lo siguiente sería ir cogiendo algunas moléculas del set de 30 (las que sean fáciles de entender, con un SMILES corto) y comenzar con el método de canonización.
* Para lo anterior, necesitaré entender y estudiar el código de openbabel, y su estructura de clases. Tendré que hacer también un diagrama de clases general (será bastante enorme) y quizás uno con resaltando las clases que yo vaya a utilizar/modificar/ampliar.

## Próxima reunión

La próxima reunión está fechada para el 01/03/23

# Tercera reunión (01/03/23)

## Objetivos de la reunión

* Revisión del sprint hasta el momento
* Solucionar algunas dudas

## Preguntas para esta reunión

* No son 100 moléculas, eran 30 los del excel. No sé si afectará mucho esto.
  + Respuesta: no importa, usamos este conjunto de datos.
* Preguntar sobre este documento de las actas, si lo ve bien.
  + Respuesta: sí, está bien. Luego esto lo puedo anexar como informes de sprints.
* En google colab: Preguntar por los distintos modos de dibujado ¿Qué paquete está dibujando cada molécula?
  + Respuesta: el nombre del import Chem.Draw de Rdkit coincide con el nombre del programa (que no tiene que ver) ChemDraw, por lo que lo dibuja Rdkit propiamente. Openbabel igual, tiene sus propios métodos de dibujar.
* Porque en algunos SMILES usan el “.” Y luego en el dibujo aparecen con enlaces. ¿No se supone que eso son moléculas desconectadas? P.ej el 16 del excel. O más fácil, el 4 del excel: en sigma aldrich tiene un “.”, pero el dibujo sale junto; en pubchem sale el “.” Y el dibujo sale separado; y en SciFinder no aparece “.” Y el dibujo sale junto.
  + Respuesta: se supone que el punto indica desconexión, pero no siempre. Puedes hacer enlaces a mano usando los números de los ciclos (no se supone que tengan que servir para eso, pero se puede)
* ¿Trabajas con químicos del ICIQ? ¿Con cuántos? ¿En qué área trabajan, específicamente en organometalica? Todo esto para meter un párrafo en la motivación de la memoria, para poner en contexto
  + Respuesta: si, trabaja con un grupo de investigación de Tarragona, rocío se comunica con la jefa del grupo. Sí, en concreto de organometalica.
* ¿Hasta qué punto les supone un problema a los investigadores/químicos que usan estas bases de datos, estas diferencias entre representaciones?
  + Respuesta: Su problema como tal no es que sean distintas, sino que no pueden dibujarlos, de ninguna de las maneras de las bases de datos. Por lo visto el código SMILES no es muy apto para organometálica, les dificulta el trabajar con estas moléculas, si no las pueden dibujar, no las pueden entender. El problema no es tanto de la estructura de la cadena, sino de que luego puedan trabajar con ella. Lo que necesitan ellos, es que SMILES codifique correctamente una molécula de organometalica para poder luego usar todas las herramientas. Hasta el momento, según comenta rocío, no hay ninguna notación que haga esto bien. De hecho, no tiene por qué ser SMILES, la cosa es obtener una nomenclatura que les sirva. Nos centramos en SMILES porque es la más usada e intentamos “parchearla”
* Preguntar por la estructura de la memoria: ¿estado del arte?
  + Respuesta: habrá que hacer algo si. Podré meter una revisión bibliográfica de la literatura existente (tampoco muy extenso esto). Tendré que hablar por aquí de los distintos paquetes que existen (rdkit, openbabel); las distintas formas de representar una molécula (SMILES, InChI, SELFIES) con una imagen y una explicación corta/media de cada una.

Los fundamentos teóricos sin pasarme tampoco, que sino se me hace muy largo este capitulo

* En la tabla comparativa inicial: yo lo he enfocado más entre Scifinder y sigma-aldrich ¿incluyo también PubChem? ¿inlcuyo tb los dibujos que yo mismo pueda generar con rdkit y openbabel?
  + Respuesta: con esas 2 va bien, en la motivación es solamente para mostrar que cada fuente de datos hace más o menos lo que quiere. Puedo hacer una tabla más pequeña como la que tengo para la motivación, a modo de ejemplo. Luego, con el dataset de moléculas, hacer una tabla más grande (en horizontal si es necesario, como anexo/apéndice).

## Conclusiones de la reunión

- Han sido todo prácticamente preguntas, lo importante ya lo he anotado

- Terminar los apartados por completar del primer capítulo

## Tareas nuevas que hacer

* Hacer diagrama de gantt de planificación, aunque sea grosso modo, para ver los tiempos que llevo y ver si voy lento o no.

## Tareas pendientes

* Hacer una tabla resumen con las diferencias entre los códigos SMILES y dibujos para una misma molécula en las distintas webs. Usar también Rdkit, hay algunas moléculas que no soporta. Esto para la introducción/motivación. 🡪 **Por completar**
* Escribir la introducción al trabajo entorno a lo anterior (química informática, sistemas de representación, organometalica…) 🡪 **Por completar**

Menos prioritario:

* Estudiar metodología de trabajo (tradicional o agil) 🡪 **Por justificar**
* Recursos: que costo tiene mi proyecto 🡪 **Por hacer**
* Analisis de riesgos, planes de contingencia 🡪 **Por hacer**

## Tareas a futuro

Sin cambios respecto a la anterior reunión.

## Próxima reunión

La próxima reunión está fechada para el 16/03/2023