



TRABAJO FIN DE GRADO
INGENIERÍA EN ...

Titulo del Proyecto

Subtitulo del Proyecto

Autor

Nombre Apellido1 Apellido2 (alumno)

Directores

Nombre Apellido1 Apellido2 (tutor1)

Nombre Apellido1 Apellido2 (tutor2)



ESCUELA TÉCNICA SUPERIOR DE INGENIERÍAS INFORMÁTICA Y DE
TELECOMUNICACIÓN

Granada, mes de 201



Título del proyecto

Subtítulo del proyecto.

Autor

Nombre Apellido1 Apellido2 (alumno)

Directores

Nombre Apellido1 Apellido2 (tutor1)

Nombre Apellido1 Apellido2 (tutor2)

Título del Proyecto: Subtítulo del proyecto

Nombre Apellido1 Apellido2 (alumno)

Palabras clave: palabra_clave1, palabra_clave2, palabra_clave3,

Resumen

Poner aquí el resumen.

Project Title: Project Subtitle

First name, Family name (student)

Keywords: Keyword1, Keyword2, Keyword3,

Abstract

Write here the abstract in English.

Yo, **Nombre Apellido1 Apellido2**, alumno de la titulación TITULACIÓN de la **Escuela Técnica Superior de Ingenierías Informática y de Telecomunicación de la Universidad de Granada**, con DNI XXXXXXXXXX, autorizo la ubicación de la siguiente copia de mi Trabajo Fin de Grado en la biblioteca del centro para que pueda ser consultada por las personas que lo deseen.

Fdo: Nombre Apellido1 Apellido2

Granada a X de mes de 201 .

D. **Nombre Apellido1 Apellido2 (tutor1)**, Profesor del Área de XXXX del Departamento YYYY de la Universidad de Granada.

D. **Nombre Apellido1 Apellido2 (tutor2)**, Profesor del Área de XXXX del Departamento YYYY de la Universidad de Granada.

Informan:

Que el presente trabajo, titulado ***Título del proyecto, Subtítulo del proyecto***, ha sido realizado bajo su supervisión por **Nombre Apellido1 Apellido2 (alumno)**, y autorizamos la defensa de dicho trabajo ante el tribunal que corresponda.

Y para que conste, expiden y firman el presente informe en Granada a X de mes de 201 .

Los directores:

Nombre Apellido1 Apellido2 (tutor1) **Nombre Apellido1 Apellido2 (tutor2)**

Agradecimientos

Poner aquí agradecimientos...

Índice general

Índice de figuras

1.	Estructura de la cafeína	2
2.	Estructura del metano	3
3.	Dos formas de representar el butano	3
4.	Fases de la metodología SCRUM [2]	7
5.	Dos posibles codificaciones de la melatonina en SMILES [16]	10
6.	Ciclos en SMILES	10
7.	Contenido de un archivo MOL [6]	11

Índice de cuadros

Introducción

Debido al desarrollo exponencial de la informática en el último siglo, todas las ramas del conocimiento se han visto afectadas. La informática está permitiendo la automatización de muchas tareas que anteriormente se realizaban de forma manual por un operario.

Con el paso de los años, esta capacidad de automatización va en aumento pudiéndose aplicar en situaciones anteriormente impensables. Los numerosos avances en hardware y la aplicación de arquitecturas GPU en el campo del Aprendizaje Automático, y concretamente del Deep Learning, han permitido crear modelos mucho más avanzados capaces de interiorizar datos más complejos.

La química es una de las ciencias que se ha impregnado de este desarrollo tecnológico y de esta combinación ha surgido lo que se conoce como cheminformatics o chemoinformatics. En este capítulo definiremos esta rama científica y describiremos sus principales actuaciones. A continuación, explicaremos distintos conceptos y tecnologías existentes relacionados con el proyecto.

Motivación del proyecto

Desde hace décadas, en el mundo de la química ha estado presente la necesidad de almacenar, gestionar y procesar la gran cantidad de información que se genera. Con el tiempo se fueron desarrollando técnicas de tratamiento de ésta, pero no fue hasta hace algunos años cuando se acuñó el nombre de cheminformatics o chemoinformatics.

En la literatura existen diferentes definiciones para este término, discutidas en [3].

“Chem(o)informatics es un término genérico que engloba el diseño, creación, gestión, recuperación, análisis, diseminación, visualización y el uso de información química” es una de las definiciones recogidas. Otra más abierta es “La aplicación de métodos informáticos para resolver problemas de química”.

Desde la Universidad de Granada, mi tutora Rocío trabaja en este ámbito. Colabora con químicos de la Universidad de Negev y es consciente de los problemas que tienen para manejar la gran cantidad de datos que aparecen en publicaciones científicas. Un tipo de datos muy valioso son las imágenes, pero clasificarlas no es trivial: pueden ser sobre cualquier temática, algunas pueden referirse a esquemas explicando cómo funciona un modelo, otras pueden contener resultados de algún experimento, pueden ser representaciones de compuestos químicos, etc. Clasificarlas manualmente por un operario no es una opción viable.

Es por ello que en este Trabajo de Fin de Grado vamos a crear una utilidad que permita clasificar imágenes. En concreto, aquellas que son representaciones de moléculas organometálicas, del resto.

Aunque existen diferentes opiniones sobre el alcance de las cheminformatics, se puede considerar que este proyecto está dentro de sus fronteras, ya que vamos a crear una herramienta de clasificación de imágenes químicas, es decir, una herramienta que procesa y analiza este tipo de información.

Compuestos orgánicos y su representación

Un compuesto orgánico es un compuesto químico que contiene átomos de carbono, formando enlaces carbono-carbono y carbono-hidrógeno [12]. En este TFG, vamos a trabajar concretamente con la clasificación de compuestos organometálicos, donde los átomos de carbono forman enlaces covalentes con átomos metálicos [13].

En las publicaciones de química encontramos numerosas representaciones gráficas de estos compuestos, lo que se conocen como fórmulas estructurales. Éstas muestran la disposición en el espacio de los átomos que forman el compuesto. Un ejemplo es la siguiente figura:

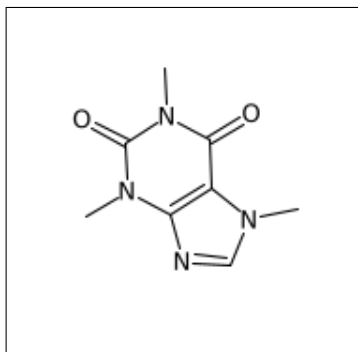


Figura 1: Estructura de la cafeína

En ellas pueden aparecer multitud de elementos, apareciendo siempre:

- **Átomos:** Se sitúan en los extremos de los enlaces. Representados con letras que indican el elemento químico del que se trata, tal y como aparece en la tabla periódica.
- **Enlaces:** Unen dos átomos entre sí.

En algunas ocasiones, sobre los átomos pueden mostrarse cargas positivas o negativas representando iones. Aparte, puede mostrarse lo que se conoce como información estereoquímica, que indica la disposición de los átomos en el espacio. Ésta es importante ya que afecta a las propiedades y reactividad de las moléculas [14, 10].

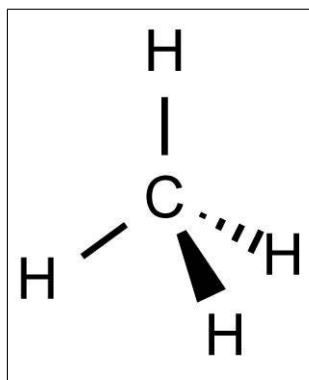


Figura 2: Estructura del metano

- Las líneas sólidas representan enlaces en el plano.
- Las discontinuas representan enlaces que están más alejados.
- Aquellas con forma de cuña indican que uno de los átomos se encuentra más cerca del espectador.

Además, un mismo compuesto se puede representar de distintas formas:

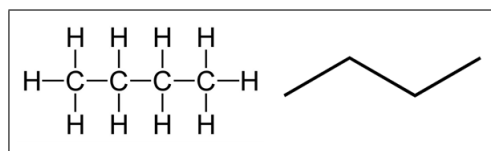


Figura 3: Dos formas de representar el butano

A la izquierda se detalla la definición de todos los átomos, en cambio a la derecha encontramos lo que se conoce como fórmula de esqueleto, donde

se omiten los átomos de carbono e hidrógeno. Se sabe que hay un átomo de carbono en los vértices que quedan libres en la intersección de dos enlaces o en las terminaciones donde no aparece ningún otro elemento. Se supone a la vez que cada átomo de carbono tiene cuatro enlaces, por tanto el número de enlaces que faltan por indicar explícitamente se corresponden con enlaces a moléculas de hidrógeno [15, 10].

Planificación del proyecto

En los primeros años del desarrollo de software, este se creaba sin seguir ningún enfoque formal. Muchos de los proyectos que se iniciaban terminaban fracasando por los retrasos en la entrega, el mal funcionamiento del producto o por no cumplir los requisitos del cliente. Además, la complejidad requerida en el software iba aumentando con el tiempo.

Por ello, era necesario crear un marco que permitiera metodizar el desarrollo. Es así como surge la Ingeniería del Software, que acompaña al software durante todo su ciclo vital.

En nuestro proyecto vamos a aplicar este enfoque a la hora de desarrollar el producto, y en este apartado vamos a describir información general sobre su gestión, la metodología de desarrollo a seguir y otra información relativa a costes, riesgos y otros factores que influyen sobre el proyecto para asegurarnos de que se consiguen los objetivos en el tiempo previsto, detectando posibles amenazas y problemas a tiempo.

Gestión del proyecto

Metodología de desarrollo

Las metodologías se pueden clasificar en dos grandes bloques [9], tradicionales y ágiles. Las tradicionales son las que primero surgieron, se caracterizan por definir rígidamente los requisitos al inicio del proyecto. En ellas, se aplica una serie de etapas de forma lineal, y una vez alcanzada una de ellas no se puede volver atrás. Por todo esto no se adaptan bien a los cambios.

En general, los proyectos de software tienden a ser cada vez más complejos. Las metodologías ágiles surgieron con el objetivo de hacer los proyectos de desarrollo más dinámicos, de forma que se adaptaran mejor al entorno y a los cambios. Se basan en una metodología incremental, donde se van construyendo prototipos del producto poco a poco, añadiendo funcionalidades hasta obtener la aplicación final. Los equipos se reúnen cada poco tiempo

para intercambiar ideas y repartir las tareas a realizar.

A la hora de elegir la metodología que vamos a usar, debemos tener en cuenta los siguientes factores relativos a la naturaleza del proyecto:

- Se trata de un proyecto de investigación, donde vamos a aplicar diferentes técnicas de Aprendizaje Automático a la resolución de un problema. Por tanto, a priori no se conoce la calidad de los resultados que se van a obtener y el número y tipo de experimentos que va a ser necesario realizar.
- La intervención del experto en química va a ser fundamental durante el desarrollo del proyecto. Aportará información y feedback esencial durante todas las fases.

Por estas razones, creemos que la metodología de desarrollo que mejor se adapta a nuestras necesidades es una metodología ágil, ya que nos provee de gran flexibilidad, permitiendo el desarrollo del proyecto de una forma incremental donde obtenemos feedback del cliente y del experto en cada iteración.

Revisando las distintas metodologías [4], creemos que una buena candidata es SCRUM [2]. Es posiblemente una de las más utilizadas en la actualidad, y los proyectos que la aplican cuentan con las siguientes características:

- **Entregable flexible:** Su contenido viene dado por lo que demanda el entorno.
- **Calendario flexible:** El entregable puede ser requerido antes o después de lo previsto.
- **Equipos pequeños:** Los equipos están formados por pocas personas, de forma que la comunicación y sincronización entre sus miembros es alta.
- **Revisiones frecuentes:** El progreso del equipo se evalúa de forma periódica y frecuentemente, de forma que se ponen en común las dificultades encontradas y se intentan resolver con la ayuda de todos los miembros.
- **Colaboración:** La colaboración entre todos los miembros del equipo es muy alta, así como entre el equipo y entidades externas como el cliente.

Cuando se trabaja con esta metodología, en cada equipo existe un miembro conocido como SCRUM manager. Es el encargado de guiar al equipo en

el desarrollo y en la aplicación de la metodología. En SCRUM el equipo es muy importante y todos sus miembros participan con su opinión. Esta metodología consta de las siguientes fases:

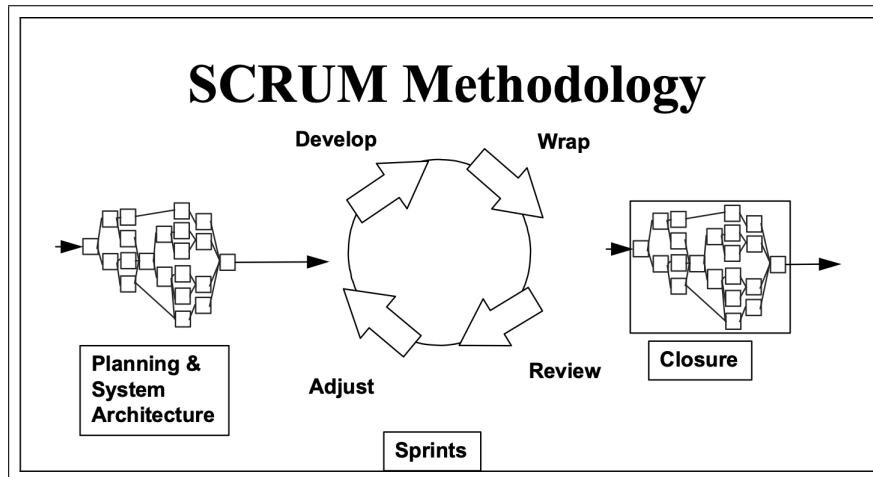


Figura 4: Fases de la metodología SCRUM [2]

■ **Pregame:**

- *Planificación:* En esta fase se crea la lista de tareas a realizar (backlog list), se fija la fecha de entrega del producto y su funcionalidad, se forma el equipo (o equipos) de trabajo, se valoran los riesgos que puedan surgir, los costes y finalmente se revisa todo y se aprueba el proyecto.
- *Arquitectura/Diseño de alto nivel:* Se revisan los elementos en el backlog, se realizan cambios si es necesario para poder implementarlos, se perfila la arquitectura del sistema teniendo en cuenta estos cambios y una reunión es organizada para revisar el diseño, donde cada equipo presenta una propuesta para implementar cada backlog.

■ **Game:** Corresponde con el conocido Sprint de la metodología SCRUM. Es la parte iterativa de la metodología, que se ejecuta varias veces hasta conseguir el producto final y está formada por las siguientes subtareas:

- *Desarrollo:* Lo primero que se realiza es definir los cambios que hay que realizar en los backlogs para poder implementarlos. Se dividen las tareas presentes en el backlog en paquetes, y se completan estos paquetes diseñando, desarrollando, implementando, testeando y documentando los cambios.

- *Envoltura*: Se cierran los paquetes, creando una ejecutable que incorpora los cambios y se explica como cumplen lo especificado en los backlogs.
 - *Revisión*: Todos los equipos se reúnen para presentar el trabajo. El desarrollo obtenido se evalúa y se añaden nuevas tareas que puedan surgir al backlog. Se evalúa el riesgo y se realizan propuestas en base a este.
 - *Ajuste*: Se consolida la información recibida durante la revisión.
- **Postgame:**
- *Cierre*: Cuando el gestor del proyecto considera que el producto está terminado y cumple con los requisitos solicitados por el cliente se entra en esta fase, donde se prepara el producto para su despliegue. Integración, generación de la documentación, testeo y marketing son algunas de las actividades que se realizan en este paso.

Estas características que hemos mencionado son las especificadas en la definición original de SCRUM. Aún así, en la práctica cada proyecto las adapta a sus necesidades. En general, los sprints suelen tener una duración de 2 o 3 semanas. En nuestro caso tendrán 1 o 2 semanas de duración.

Estado del arte

Para poder organizar este proyecto, debemos conocer el estado del arte en su ámbito. Como comentamos en la introducción, cheminformatics es un tema muy amplio y engloba subtópicos diferentes. Muchos de ellos surgieron en la década de 1960 y principios de 1970, y desde esa época muchos grupos de investigación siguen trabajando en ellos y nuevos grupos han surgido para aplicar nuevas tecnologías en nuevos ámbitos. A continuación discutimos algunos de los temas que históricamente han sido de interés para esta ciencia. [3]

Representación de compuestos y reacciones químicas

El poder almacenar representaciones de compuestos químicos de manera eficiente en un ordenador requiere de la creación de métodos y formatos específicos para ello. Además, hay que tener en cuenta qué datos vamos a codificar, si solo la estructura básica del compuesto, si también queremos guardar información estereoquímica o si queremos añadir notas auxiliares sobre los compuestos. Es importante tener esto claro, ya que la complejidad de la representación influirá en la cantidad de almacenamiento que ocupe en disco y en los recursos necesarios para procesarla.

Utilizando notación lineal, representamos la estructura del compuesto como una secuencia lineal de caracteres y números. Esta es una codificación adecuada para las computadoras, ya que la pueden procesar con facilidad. Algunos formatos que utilizan esta notación son WLN (Wiswesser Line Notation), ROSDAL (Rp) o SMILES (Simplified Molecular Input Line Entry Specification). Aunque WLN y ROSDAL han quedado obsoletos, SMILES se sigue utilizando con mucha frecuencia en la actualidad. [3]

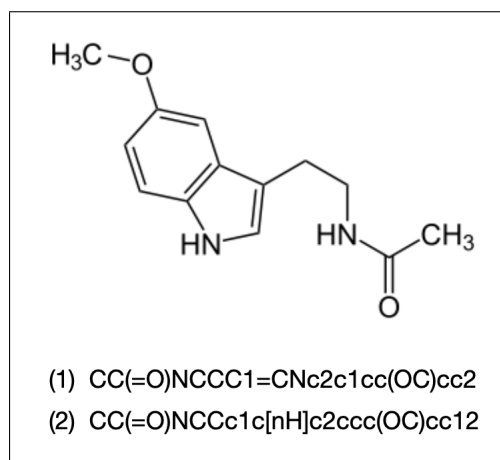


Figura 5: Dos posibles codificaciones de la melatonina en SMILES [16]

Aunque no vamos a entrar en detalle de cómo funciona, ya que para ello hay que tener nociones avanzadas de química, diremos que utiliza las siglas de cada elemento de la tabla periódica para representar los átomos. La primera letra del elemento se escribe en mayúscula, a no ser que se trate de un átomo perteneciente a un anillo aromático¹, ya que en ese caso se escribe en minúscula. Si el elemento tiene dos caracteres, el segundo se escribe siempre en minúscula. Además, se pueden representar cargas.

Los enlaces se representan con -, =, # y :, según el tipo. Bajo algunas circunstancias, se pueden omitir estos símbolos, ya que por su contexto se deducen. También se pueden codificar ramas, situando elementos entre paréntesis, y ciclos, utilizando un número para indicar el inicio y el fin del ciclo en la cadena de texto. [1]

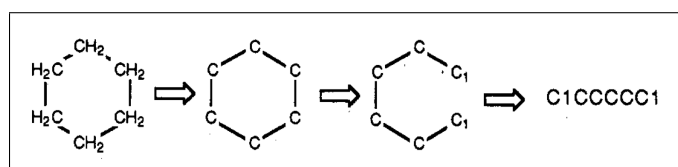


Figura 6: Ciclos en SMILES

Otras características como la aromaticidad o estructuras inconectas también pueden ser representadas. Entre las ventajas de esta codificación, destaca su facilidad de comprensión por los humanos. Cualquier químico puede aprender sus reglas de codificación fácilmente y diseñar sus propios com-

¹La aromaticidad es una propiedad presente en enlaces dobles de moléculas cíclicas, donde sus electrones pueden circular libremente. Esto mejora la estabilidad del compuesto. [7, 11]

puestos. Un problema que tiene SMILES es que un mismo elemento se puede representar de diferentes formas. Pero sobre todo, el mayor problema es que un porcentaje significativo de las cadenas no se corresponden con moléculas válidas, ya sea porque son sintácticamente inválidas, no se corresponden con un grafo molecular o no cumplen reglas químicas básicas. [1]

Uno de los principales objetivos de la química computacional es el diseño de nuevas moléculas. Para ello, la utilización de modelos generativos puede ayudar a los investigadores, pero si el espacio de estados de SMILES no es completamente válido se dificulta la tarea. Para ello han surgido otras codificaciones como SELFIES con un espacio 100 % robusto. [5]

Además de estos formatos de notación lineal, es necesario mencionar otros. El lanzamiento en 1982 de MDL Molfile llevó a su aceptación como principal formato para representar datasets químicos. Se han realizado distintas adaptaciones de este para añadir información extra a las moléculas, dando lugar a SDfile, RGfile, Rxnfile, etc. El formato PDB se utiliza principalmente para almacenar información 3D de macromoléculas biológicas, como son las proteínas o los polinucleótidos. CIF también es un formato para almacenar información 3D. En espectroscopia encontramos JCAMP. Finalmente, CML (chemical markup language), una extensión de XML, es una propuesta que intenta aglutinar toda la información disponible. Es compatible con moléculas, reacciones, espectroscopia y otra información. [3]

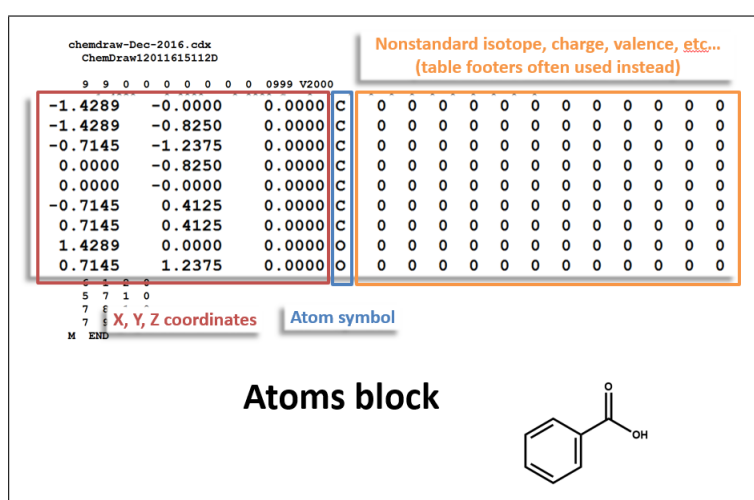


Figura 7: Contenido de un archivo MOL [6]

Fuentes y bases de datos

El gran número de facetas que presenta la información química necesita sistemas de almacenamiento a la altura. La química fue una de las primeras ramas científicas en utilizar bases de datos para almacenar: la cantidad de datos que se generaba creció rápidamente, y sigue creciendo hoy en día.

Aunque es muy complicado clasificarlas, vamos a separarlas en tres grandes grupos según el tipo de información que almacenan: [3]

- **Bases de datos de publicaciones:** Pueden ser bibliográficas, guardando solamente los metadatos y la referencia a la publicación, o de texto completo, donde recogen la publicación de forma íntegra.
- **Bases de datos fácticas:** Al contrario que las anteriores, que almacenan publicaciones de la literatura primaria, estas pueden guardar propiedades físicas de compuestos, información de espectroscopia, información legal, etc.
- **Bases de datos de estructuras y reacciones:** Recogen estructuras químicas, tanto individualmente como formando parte de reacciones. No se almacenan como imágenes, sino en formatos interpretables por la máquina.
- **Bases de datos de biología molecular:** Contienen secuencias de aminoácidos y nucleótidos.

Pero, ¿cómo podemos rellenarlas con datos? ¿De dónde podemos extraerlos? Durante décadas se han publicado un gran número de artículos. Podríamos utilizarlos como una fuente muy amplia de información, ya que contienen todo el progreso científico. Específicamente para extraer datos relativos a estructuras, entran en juego utilidades conocidas como OCSR (Optical Chemical Structure Recognition).

Son capaces de transformar una imagen en un formato compatible con la máquina, como podría ser SMILES. En muchos casos incluso es posible introducirles una publicación completa y ellas mismas localizan las imágenes de moléculas. Con el paso del tiempo se han ido perfeccionando, y algunas son capaces de detectar información estereoquímica o de relacionar el compuesto con el texto de la publicación. Más adelante repasaremos algunas de estas utilidades.

Por último, mencionar una base de datos que merece la pena conocer. Creada por el National Institute of Health (NIH), PubChem es una base de datos abierta que cada mes sirve a millones de usuarios en todo el mundo. Es una base de datos de estructuras y aunque contiene mayoritariamente moléculas pequeñas, también almacena nucleótidos, carbohidratos, lípidos,

péptidos y macromoléculas modificadas químicamente. Para cada compuesto almacena su estructura, identificadores, propiedades físicas y químicas, toxicidad, patentes, etc. Los datos que aglutina provienen de diversas fuentes, como son agencias del gobierno estadounidense, editores de revistas científicas o proveedores químicos, aunque hay muchas más. [8]

Métodos de búsqueda

Almacenar información en las bases de datos no sirve de nada si no se desarrollan métodos eficientes para extraerla. En aquellas bases de datos donde se almacenan estructuras químicas, una de las principales formas de obtener información es buscar similitudes entre una molécula dada como entrada y otras que se encuentran almacenadas, de forma que compartan una subestructura específica o tengan otras características en común. Para ello, es clave la codificación de los compuestos.

También, si se almacenan metadatos y la base de datos está indexada sobre ellos se podría buscar por su nombre, etiquetas, etc. [3]

Métodos para análisis de datos

En química, grandes cantidades de datos son producidas. Una vez que hemos conseguido limpiarlos y ordenarlos, tenemos un conocimiento muy valioso en nuestras manos. La información es muy interesante en sí misma, pero también lo son las relaciones que se esconden en su interior. Para ello, se crean modelos que puedan interiorizarlas.

El análisis de datos no solo se enfrenta a la extracción de la información principal, sino que también intenta generar nueva información secundaria. [3]

Este TFG se desarrolla dentro de este ámbito, ya que, como describiremos más adelante, entrenamos un modelo capaz de detectar que imágenes contienen moléculas organometálicas. Además, creamos un modelo generativo para aumentar el tamaño del dataset en uso.

Bibliografía

- [1] David Weininger. “SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules”. *Journal of chemical information and computer sciences* 28.1 (1988), págs. 31-36.
- [2] Ken Schwaber. “Scrum development process”. *Business object design and implementation*. Springer, 1997, págs. 117-134.
- [3] Thomas Engel. “Basic Overview of Chemoinformatics”. *Journal of Chemical Information and Modeling* 46.6 (2006). PMID: 17125169, págs. 2267-2277. DOI: 10.1021/ci600234z. eprint: <https://doi.org/10.1021/ci600234z>. URL: <https://doi.org/10.1021/ci600234z>.
- [4] Mihai Liviu Despa. “Comparative study on software development methodologies”. *Database Systems Journal* 5.3 (2014), págs. 37-56.
- [5] Mario Krenn y col. “Self-referencing embedded strings (SELFIES): A 100 % robust molecular string representation”. *Machine Learning: Science and Technology* 1.4 (oct. de 2020), pág. 045024. DOI: 10.1088/2632-2153/aba947. URL: <https://doi.org/10.1088/2632-2153/aba947>.

Otras fuentes

- [6] Robert Belford. *Anatomy of a MOL file*. URL: [https://chem.libretexts.org/Courses/University_of_Arkansas_Little_Rock/ChemInformatics_\(2017\)%3A_Chem_4399_5399/2.2%3A_Chemical_Representations_on_Computer%3A_Part_II/2.2.2%3A_Anatomy_of_a_MOL_file](https://chem.libretexts.org/Courses/University_of_Arkansas_Little_Rock/ChemInformatics_(2017)%3A_Chem_4399_5399/2.2%3A_Chemical_Representations_on_Computer%3A_Part_II/2.2.2%3A_Anatomy_of_a_MOL_file) (visitado 26-04-2022).
- [7] curiosoando.com. *¿Qué es la aromaticidad?* URL: <https://curiosoando.com/que-es-la-aromaticidad> (visitado 25-04-2022).
- [8] National Institute of Health (NIH). *PubChem Docs - About PubChem*. URL: <https://pubchemdocs.ncbi.nlm.nih.gov/about> (visitado 26-04-2022).
- [9] Santander Universidades. *Metodologías de desarrollo de software: ¿qué son?* URL: <https://www.becas-santander.com/es/blog/metodologias-desarrollo-software.html> (visitado 04-03-2022).
- [10] Aditya Virani y col. *Structural Representations of Organic Compounds*. URL: <https://brilliant.org/wiki/structural-representations-of-organic-compounds/> (visitado 21-04-2022).
- [11] Wikipedia. *Aromaticidad*. URL: <https://es.wikipedia.org/wiki/Aromaticidad> (visitado 25-04-2022).
- [12] Wikipedia. *Compuesto orgánico*. URL: https://es.wikipedia.org/wiki/Compuesto_org%C3%A1nico (visitado 21-04-2022).
- [13] Wikipedia. *Compuesto organometálico*. URL: https://es.wikipedia.org/wiki/Compuesto_organomet%C3%A1lico (visitado 21-04-2022).
- [14] Wikipedia. *Estereoquímica*. URL: <https://es.wikipedia.org/wiki/Estereoqu%C3%ADmica> (visitado 21-04-2022).
- [15] Wikipedia. *Fórmula estructural*. URL: https://es.wikipedia.org/wiki/F%C3%B3rmula_estructural (visitado 21-04-2022).
- [16] Wikipedia. *Simplified molecular-input line-entry system*. URL: https://en.wikipedia.org/wiki/Simplified_molecular-input_line-entry_system (visitado 24-04-2022).