Міністерство освіти і науки України
Донбаська державна машинобудівна академія (ДДМА)

КОМП'ЮТЕРНІ ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ В ХІМІЇ

Стислий конспект лекцій

для студентів спеціальності 102 «Хімія»
денної форми навчання

Затверджено
на засіданні
методичної ради
Протокол № від

Краматорськ
ДДМА
2020

УДК 543 + 544

Комп’ютерні та інформаційні технології в хімії: стислий конспект лекцій для студентів спеціальності 102 «Хімія» денної форми навчання / уклад. С. О. Коновалова. – Краматорськ : ДДМА, 2020. – 80 с.

Посібник містить стислий конспект лекцій з дисципліни «Комп’ютерні та інформаційні технології в хімії». В посібнику наведено основні підходи до представлення хімічної структури від лінійних нотацій до представлення молекулярних поверхонь, наведені основні характеристики сучасних редакторів хімічних формул та програмних засобів, що використовуються для візуалізації молекулярних моделей і структурної інформації усіх рівнів. Окремий розділ охоплює проблеми пошуку хімічної інформації в онлайнових джерелах і доступних базах хімічних даних. Посібник складено з метою зменшення непродуктивних витрат часу студента на підготовку до занять та сприяє більш раціональному плануванню часу.

Укладач С. О. Коновалова, доц.

Відп. за випуск А. П. Авдєєнко, проф.

ЗМІСТ

ВСТУП 4
1 КОМП'ЮТЕРНІ ТЕХНОЛОГІЇ В ХІМІЇ 6
1.1 Програми для візуалізації хімічних структур 6
1.1.1 Ліцензії на програмне забезпечення 6
1.1.2 Огляд програмного забезпечення для створення хімічних формул 9
1.2 Створення молекулярних моделей за допомогою спеціального програмного забезпечення 11
1.2.1 1-D Рівень представлення хімічних структур 11
1.2.2 2-D Рівень представлення хімічних структур 15
1.2.3 3-D Рівень представлення хімічних структур 26
1.3 Інтеграція даних спеціалізованих програм до програм пакету MS Office 35
2 ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ В ХІМІЇ 41
2.1 Бази даних та хімічні каталоги в мережі Internet 41
2.1.1 Ресурс PubChem 42
2.1.2 Ресурс ChemSpider 45
2.1.3 Ресурс e-Molecules 47
2.1.4 Ресурс COMMON CHEMISTRY (CAS) 49
2.1.5 Ресурс Organic Syntheses 50
2.2 Електронні ресурси хімічної наукової періодики 54
2.2.1 Наукова періодика України 54
2.2.2 Наукова періодика світу 57
2.2.3 Патентні бази даних 63
2.2.4 Системи ідентифікації науковців і оцінювання наукової діяльності 64
2.2.5 Основи роботи в міжнародній науковій базі Scopus 68
2.2.6 Пошук інформації з хімії у мережі Internet 76
ЛІТЕРАТУРА 79

ВСТУП

Використання інформаційних комп'ютерних технологій в хімії тісно пов'язано з розвитком хімічної інформатики. Хімічна інформатика – це загальний термін, що поєднує в собі дизайн, створення, організацію, керування, перетворення, аналіз, поширення, візуалізацію і використання хімічної інформації.

Перетворення даних в інформацію, а інформації в знання є необхідним процесом в будь-якій хімічній дисципліні. Розвиток хімічної інформатики регулярно розширює межі використання комп'ютерних технологій, методів статистики, методів візуалізації для вирішення хімічних задач.

Хімія призводить до утворення значної кількості даних, яка зростає лавино подібно. Вже зараз відомо понад 70 мільйонів хімічних сполук, і їх кількість зростає кожного року. Комбінаторна хімія і скрінінг генерують величезні кількості нових даних, які можна раціонально використовувати лише за умови їх зберігання у вигляді спеціалізованих баз даних.

З іншого боку, для вирішення багатьох хімічних задач необхідна інформація відсутня. Інформація про 3D-будову, отримана методом рентгеноструктурного аналізу, на даний час відома лише для 300000 органічних сполук, а найбільша база даних ІЧ-спектрів містить інформацію про 200000 сполук. Ця кількість інформації здається надзвичайно великою, але це всього лише мала доля від кількості відомих на сьогодні сполук. Тому постає питання прогнозування необхідної інформації для нових сполук на базі вже відомих даних.

Усі ці хімічні задачі вимагають нових підходів до їх вирішення, і саме в цій сфері потрібні методи хімічної інформатики, які успішно використовуються у різних областях хімічної науки: фізичній, аналітичній, органічній хімії, біохімії, матеріалознавстві, хімії полімерів тощо.

Невпинне зростання кількості хімічних сполук, інформація про які зберігається у корпоративних, комерційних та відкритих базах даних, піднімає питання про взаємозв'язок наявних бібліотек, аналіз і візуалізацію хімічного простору. Поняття структурного дескриптору і хімічного простору є ключовими для всіх сфер застосування хімічної інформатики. ***Структурний дескриптор*** – незалежна змінна, яка характеризує структурні особливості хімічної сполуки або її фрагмента. ***Хімічний простір*** – загальний простір дескрипторів, який охоплює усі невеликі органічні молекули, які можуть бути принципово створені. Для аналізу і візуалізації такого хімічного простору розроблено цілу низку методів, які успішно застосовуються для порівняння хімічного простору колекцій хімічних сполук з різного спрямування: природні речовини, ліки, метаболіти, токсичні та нетоксичні речовини, тощо.

Метою даного посібника є надання студентам необхідної інформації для набуття теоретичних основ і практичних навичок роботи з комп'ютером, спеціалізованими інформаційними базами даних, використання сучасного програмного забезпечення для автоматизації професійної діяльності.

У даному конспекті лекцій викладено основні підходи до представлення

хімічної структури від лінійних нотацій до представлення молекулярних поверхонь, наведені основні характеристики сучасних редакторів хімічних формул та програмних засобів, що використовуються для візуалізації молекулярних моделей і структурної інформації усіх рівнів. Окремий розділ охоплює проблеми пошуку хімічної інформації в онлайнових джерелах і доступних базах хімічних даних.

Курс лекцій містить два розділи «Комп'ютерні технології в хімії» та «Інформаційні технології в хімії», але цей розподіл є досить умовним, тому що на даному етапі розвитку технологій їх неможливо відокремити один від одного. Тому в кожному з цих розділів наявна інформація, що стосується як комп'ютерних так і інформаційних технологій.

1 КОМП'ЮТЕРНІ ТЕХНОЛОГІЇ В ХІМІЇ

На даний час комп'ютери одержали широке поширення у всіх сферах життєдіяльності людини. Комп'ютери використовуються як засіб комунікації, як домашній центр розваг, для ведення ділової та технічної документації, конструювання в машинобудуванні та архітектури, в медицині, авіації, автомобілях, побутовій техніці, економіці, науці і всіх галузях промисловості. Особливо велике значення комп'ютери грають у хімії, зокрема, у хімічних лабораторіях і в галузях промисловості, які потребують контролю перебігу хімічних процесів.

Основними напрямками застосування комп'ютерів в хімії і промисловості являються:

– управління технологічними процесами і контроль якості готової продукції;

– моделювання складних хімічних реакцій та розробка нових хімічних сполук, зокрема, особливо складних високомолекулярних органічних сполук;

– візуалізація хімічних сполук для кращого розуміння їх будови і особливостей в плоскому і об'ємному вигляді;

– пошук хімічної інформації в онлайнових джерелах і базах хімічних даних.

1.1 Програми для візуалізації хімічних структур

1.1.1 Ліцензії на програмне забезпечення

Ліцензія на програмне забезпечення (**ліцензія на використання програмного забезпечення**) – це правовий інструмент, що визначає використання та розповсюдження програмного забезпечення, захищеного авторським правом.

Ліцензія, або ліцензійна угода – це спеціальна форма договору між автором програмного забезпечення (далі ПЗ) і користувачем програмного забезпечення. Така ліцензія визначає ***умови використання комп'ютерного ПЗ***. Ліцензія може надавати дозвіл робити з ним дії, які були б інакше заборонені законом про авторське право.

Наприклад, ліцензія на використання програмного забезпечення може дати дозвіл робити копії програмного забезпечення. Власник авторського права може запропонувати ліцензію на використання ПЗ односторонньо, або як частину ліцензійної угоди на використання ПЗ з іншою стороною.

Для позначення ліцензії на програмне забезпечення також використовують англійську абревіатуру (англ. **EULA** — End User License Agreement)

Існують різні типи ліцензій на програмне забезпечення:

- Commercial (комерційна) – закрите, власницьке або пропрієтарне ПЗ;

- Trial Software – ПЗ, яке працює лише певний проміжок часу;
- Non-Commercial Use – для некомерційного використання;
- Shareware – умовно-безкоштовні програми;
- Freeware – безкоштовне ПЗ;
- Open Source – відкрите ПЗ;
- Demo, Demoware – демонстраційне ПЗ;
- Adware – рекламно-орієнтовне ПЗ;
- Nagware, Begware – ПЗ із примусовим вікном діалогу;
- Public Domain – вільне ПЗ;
- Betaware – попередня (тестова) бета-версія комерційного або некомерційного ПЗ;
- Donateware, Donationware – авторське ПЗ.

Commercial – комерційне програмне забезпечення, яке продається за гроші, захищене різними законами. ***Пропрієтарне програмне забезпечення***, (від англ. *proprietary software*) – це програмне забезпечення, на яке зберігаються як немайнові, так і майнові авторські права. Отримавши або придбавши таке програмне забезпечення, користувач отримує обмежені права користування ним: може бути заборонено або закрито доступ до коду (вивчення), внесення змін, тиражування, розповсюдження та перепродаж. Програмне забезпечення вважається *власницьким*, якщо наявне хоча б одне з перелічених обмежень.

Хорошим прикладом пропрієтарної ліцензії може служити ліцензія на Microsoft Windows, яка включає великий список заборонених варіантів використання, таких як одночасна робота з системою декількох користувачів і поширення тестів її робочих характеристик.

Trial Software, Trialware – пробне (оціночне) програмне забезпечення. Обмежено часом використання або кількісними характеристиками, а іноді й функціоналом.

Shareware – умовно-безкоштовні програми. Клас комерційних програм з безкоштовним періодом використання. Для повнофункціонального використання вимагають оплати. При такій моделі розповсюдження пропонується спочатку випробувати програму в дії, а потім оплатити її. Щоправда, "умовність" може лежати в дуже широких межах, від простого нагадування про необхідність заплатити за програму при кожному запуску до обмеженого терміну роботи і навіть блокування в неоплаченій версії найважливіших функцій, що робить неможливим використання програми за прямим призначенням.

Freeware – безкоштовні програми. Програми без обмеження на (некомерційне) використання. Охороняються авторським правом. Слід пам'ятати, що відсутність ціни ще не означає, що виробник дозволяє її вільно поширювати, він може це і забороняти. І буває, що якась програма безкоштовна тільки для домашнього, некомерційного використання, а при використанні її в організаціях потрібно заплатити.

Open Source – відкриті програми з відкритим кодом. Можуть накладатися обмеження на модифікацію та використання в комерційних цілях. Код таких програм доступний для наступних дій:

- перегляд і вивчення;
- за наявності дозволу ліцензії – зміни, що дозволяє користувачеві узяти участь у доопрацюванні відкритої програми;
- використовувати код для створення нових програм – через запозичення коду, якщо це дозволяє сумісність ліцензій;
- виправляти в ньому помилки;
- вивчення алгоритмів, структур даних, технологій, методик та інтерфейсів (оскільки код може істотно доповнювати документацію, а за відсутності такої сам служить документацією).

Demo, Demoware – демонстраційні програми. Мають велику кількість обмежень. Основна мета – не пробне використання, а демонстрація можливостей. Помітно більш обмежене порівняно з ***Trialware***.

Adware – рекламно-орієнтовані програми. Без обмежень функціональності, але з примусовим показом реклами, яка може підвантажуватися через Інтернет без відома користувача. Зазвичай включають модуль фонового завантаження реклами, що таїть у собі небезпеку несанкціонованого дистанційного контролю комп'ютера. Антивірусні програми часто класифікують даний механізм як "троянського коня".

Nagware, Begware – основним обмеженням використання є примусове вікно діалогу, де повідомляється про те, що версія незареєстрована. Після оплати дане обмеження знімається. Ускладнює використання програми в пакетному режимі при автоматичній обробці інформації.

Public Domain – вільні програми. Без обмежень на модифікацію та використання. Не охороняються авторським правом.

Betaware – попередня (тестова) бета-версія комерційного або некомерційного ПЗ. Можна використовувати безкоштовно, але часто обмежується періодом тестування.

Donateware, Donationware – авторські програми. Для необов'язкової реєстрації програми потрібно сплатити пожертвування автору.

Існує ще багато видів програмного забезпечення, використання якого не обмежується або вимагає певних дій.

Linkware – автор програми просить вказувати посилання на сайті користувача, (якщо є) на свій сайт.

Registerware – для отримання та/або використання програми потрібно надати інформацію про себе (заповнити анкету).

Guiltware – різновид ***Nagware***. У програмі міститься явна згадка, що автор не отримав за неї грошей. Може й не передбачати реєстрації.

Crippleware – ключові можливості програми видалені. Немає обмежень на час використання. Після оплати надається повнофункціональна версія.

Abandonware – позаринкові програми. Як правило, це колишні комерційні програми, які з ряду причин перестають постачати на ринок. Їх поширює зазвичай власник авторських прав на безкоштовній основі, але з жорстким зобов'язанням заборони продавати і навіть без права подальшого безкоштовного тиражування.

Cardware, Postcardware – як компенсацію за надання програми автор просить надіслати йому листівку або електронний лист зі словами подяки. Ці листи використовуються авторами для реклами своїх робіт.

Liteware – "полегшений" варіант відповідної комерційної версії. Не обмежена часом використання, але обмежена функціоналом.

Hostageware – програми з функціональними, тимчасовими і кількісними обмеженнями. Розблоковуються після оплати.

Beerware – право на використання програми, а також отримання вихідних текстів в обмін на гроші.

Careware, Charityware – стягується збір на благодійні цілі, або безпосередньо автору, або за вказаною адресою.

Requestware – автор просить користувача щось зробити в обмін на використання програми (надіслати листівку або електронного листа з подякою, внести пожертви на благодійні цілі тощо). Різновиди: ***Postcardware, Careware***.

CDware – ПЗ на компакт-дисках, яке розповсюджується в рекламних цілях.

Spyware – програми-шпигуни. Несанкціоновано збирають інформацію про комп'ютер користувача і його дії. Нерідко маскуються під ***Adware***. Крім використання антивірусних програм найбільш ефективний спосіб боротьби з ними – встановлення брандмауерів.

1.1.2 Огляд програмного забезпечення для створення хімічних формул

На теперішній час існує достатнє велика кількість спеціального програмного забезпечення для створення хімічних формул та трьохмірної візуалізації сполук. Основна відмінність пакетів програмного забезпечення від різних виробників це ліцензія, під якою розширюється програмне забезпечення та його функціональні можливості.

ChemOffice – комерційний пакет програмного забезпечення ChemOffice складається з декількох програм – ChemDraw, Chem3D, ChemFinder, ChemFinder for Office. Авторське право: CambridgeSoft Corp. – http://www.cambridgesoft.com

1. ***ChemDraw*** – найбільш відома та популярна програма, створена спеціально для побудови структурних формул органічних і неорганічних речовин. Містить вбудовану таблицю Менделєєва, створює розрахункові ЯМР і ПМР спектри для різних речовин, генерує назви речовин за їх структурними формулами, містить велику базу шаблонів хімічних структур.

2. ***Chem3D*** – програма для створення і перегляду тривимірних хімічних структур, має вбудований плагін ChemDraw, який дозволяє за структурною формулою, відображеною на екрані, побудувати тривимірну хімічну структуру. Крім того, програма розраховує фізичні властивості хімічних структур різними квантово-хімічними методами (ММ3, МОРАС, метод Хюккеля та ін.).

3. ***ChemFinder*** – програма, яка дозволяє переглядати і створювати бази даних хімічних структур.

4. ***ChemFinder for Office*** – це дуже зручний інструмент для пошуку хімічних структур в документах Microsoft Office (*. rtf, *. doc, *. xls та ін.), файлах хімічних структур. Підтримує безліч найпоширеніших форматів, легко відшукає на комп'ютері будь-яку хімічну структуру.

ChemWindow – комплект програмного забезпечення містить чотири програми – ChemWin, SymApps, ChromKeeper, IRKeeper. Авторське право: Bio-Rad Laboratories – http://www.bio-rad.com

1. ***ChemWin*** – програма для створення структурних формул. Може автоматично перетворювати ациклічні карбонові ланцюги в структурні формули з підписами і навпаки. За структурною формулою ChemWin визначає молекулярну формулу, масовий склад сполуки та ін. Містить бібліотеку лабораторного обладнання і бібліотеку, в якій знаходяться хімічні структури.

2. ***SymApps*** – програма для перегляду об'ємних і кульострижневих моделей молекул. Добре відображає різні елементи симетрії – осі, площини симетрії та ін., чого немає, наприклад, в Chem3D Ultra. Програма вміє рахувати таблиці характерів, визначає точкові групи молекул.

3. ***ChromKeeper, IRKeeper*** – програми для роботи з експериментальними спектрами.

ACD/ChemSketch ***Freeware for personal or academic use*** **–** пакет програмного забезпечення для малювання хімічних структур. Авторське право: Advanced Chemistry Development, Inc. – http://www.acdlabs.com

ChemSketch – програма призначена для створення структурних хімічних формул. Має зручний інтерфейс, що складається з двох вікон: "Структура" і "Малюнок". Перше вікно – для зображення структур, друге – для малювання хімічних реакцій та різних схем. Програма містить дуже зручну таблицю радикалів, шаблони циклів, ланцюгів і функціональних груп та інші інструменти. У багатьох відносинах програма не поступається ChemDraw, може зберігати файли у форматах *.cdx і *.skc. Крім ChemSketch в пакет установки також входить ***3D Viewer*** – програма, що дозволяє створювати об'ємні моделі молекул.

MarvinSketch – один з кращих редакторів хімічних формул. Має безліч можливостей для редагування структур. Авторське право: ChemAxon Software, Inc. – http://www.chemaxon.com

За зручністю і можливостям його можна порівнювати з ChemDraw. Особливо варто відзначити можливість перетворення намальованих двовимірних структур в тривимірні. У комплект ***Marvin Beans*** також входять: ***MarvinView*** – програма для перегляду двох- і тривимірних хімічних структур, ***MarvinSpace*** – програма для 3D-візуалізації молекул і ***MolConverter*** – конвертер хімічних форматів.

ChemCraft (ShareWare) – квантово-хімічний візуалізатор. Авторське право: Grigoriy A. Andrienko – http://www.chemcraftprog.com

Даний візуалізатор розуміє безліч форматів хімічних даних, вихідні файли більшості популярних програм, таких, зокрема, як GAMESS і Gaussian;

має прекрасний "дружній" інтерфейс, велику кількість властивостей, які він може відображати – від довжин зв'язків до рендеринга молекулярних орбіталей і нормальних коливань. Безліч стилів відображення молекул дозволяють отримати зображення потрібної якості. ChemCraft дозволяє зручно редагувати геометрію молекул і готувати вхідні файли для розрахунків у програмах GAMESS і Gaussian. Дуже корисною є можливість експорту анімованих файлів у форматі *.gif.

1.2 Створення молекулярних моделей за допомогою спеціального програмного забезпечення

Введення і зображення хімічних структур у комп'ютерних програмах і базах даних потребує відповідних специфічних методів їх представлення. Представлення структури хімічної сполуки чи її окремого структурного фрагмента також використовується для організації збереження і пошуку сполук у базах даних.

Сполука може бути охарактеризована ***назвою***, ***лінійним кодуванням***, ***2D-структурною формулою*** та ***3D-молекулярною моделлю***. Ця ієрархія також відображає різні рівні представлення хімічної інформації в електронному форматі – від лінійних представлень до відображення 3D-структури у вигляді молекулярних поверхонь (рис. 1.1).

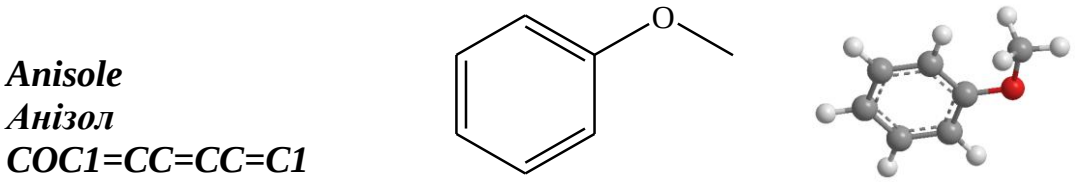


Рисунок 1.1 – Представлення хімічної структури 4-метилфенолу

1.2.1 1-D Рівень представлення хімічних структур

Лінійне кодування представляє структуру хімічної сполуки у вигляді рядка символів. Найпоширенішими є лінійні коди

– **WLN** (Wiswesser Line Notation – лінійне кодування Вісвессера);

– **ROSDAL** (Representation of Organic Structures Description Arranged Linearly – лінійне кодування органічних структур);

– **SMILES** (Simplified Molecular Input Line Entry Specification);

– **SLN** (SYBYL Line Notation).

Компактність таких лінійних записів є дуже зручною, враховуючи витрати на комп'ютерну обробку інформації. Перші лінійні коди з'явилися у 40-х роках минулого століття – WLN коди бали введені у 1946 році. Кодування ROSDAL було розроблено для Beilstein у 1985 році.

SMILES коди, розроблені у 1986 році, все ще широко розповсюджені. На відміну від WLN і ROSDAL, коди SMILES розроблені на базі декількох фундаментальних правил перетворення структури в рядок символів. Вони знайшли широке використання для представлення, розповсюдження, обміну структурною хімічною інформацією, оскільки не залежали від програмного забезпечення і внутрішньої конфігурації комп'ютерів. Сучасні правила побудови SMILES і приклади (канонічні SMILES) можна знайти в електронному ресурсі: http://www.daylight.com/dayhtml/smiles/index.html.

Основні принципи побудови кодів SMILES базуються на наступних положеннях:

– атоми кодують загальноприйнятими в хімії символами;
– атоми водню займають вільні валентності і для органічних сполук не відображаються;
– сусідні атоми наводять один за одним згідно порядку їх появи в молекулярному ланцюгу;
– подвійний і потрійний зв'язки відображаються символами «=» і «#», відповідно;
– коди бокових ланцюгів заключають в круглі дужки;
– цикли представляються шляхом присвоєння цифрових індексів двом замикаючим цикл атомам.

Деякі приклади SMILES кодування:

Ethanol	CCO
Acetic acid	CC(=O)O
Cyclohexane	C1CCCCC1
Pyridine	c1cnccc1
Sodium chloride	[Na+].[Cl-]
L-alanine	N[C@@H](C)C(=O)O

Генерування SMILES-кодів можливе декількома способами:

– власноруч відповідно до правил;
– на основі 2D-ескізів засобами редакторів хімічних формул;
– на основі 2D-ескізів засобами онлайн-ресурсів.

Власноруч можна легко записати SMILES-коди для невеликих і простих за будовою молекул. Для складних структур рекомендується використовувати засоби доступних редакторів хімічних формул або онлайн апплетів. В цьому випадку буде дотримана чітка відповідність коду правилам побудови. Такі SMILES-коди називають ще канонічними.

До програмного забезпечення, яке дозволяє генерувати SMILES-коди, відносяться редактори хімічних формул ChemDraw, ChemSketch, MarvinSketch.

Система кодів SMILES широко використовується як текстовий комунікативний формат при обміні структурними даними, оскільки вона є однозначною, і SMILES-код молекули є синонімом її структурної формули. У багатьох базах даних пошукове завдання можна формулювати у вигляді SMILES. Для пошуку інформації про хімічну сполуку у запиті код SMILES доцільно використовувати у наступних ситуаціях:

– для пошуку в текстових базах даних (де структурну формулу взагалі використати неможливо);

– в структурних базах даних, якщо код є простим у написанні (наприклад, легше набрати на клавіатурі рядок NCC(O)=O, ніж будувати двомірну структуру аланіну);

– в структурних базах даних, якщо готовий код можна скопіювати із іншого джерела, наприклад, із хімічного редактора (внаслідок несумісності форматів молекулярних редакторів не завжди можливим є перенесення двомірної формули; для символьного рядка SMILES проблема несумісності відсутня).

Розширенням кодів SMILES стали коди SMARTS для представлення хімічних карт і коди SMIRKS для представлення хімічних реакцій. Кодування ***SMARTS*** використовується для позначення субструктурних фрагментів у вигляді рядка символів і, як SMILES, буле спочатку розроблено Девідом Вайнінгером і його колегами з Daylight Chemical Information Systems,.

Поряд зі SMILES сьогодні також розповсюджені такі коди, як Sybyl Line Notation (SLN), розроблені компанією Tripos, International Chemical Identifier (InChI) і InChIKey, що розробляються в рамках наукових проектів ІЮПАК.

Шляхом модифікації SMILES фірма «Tripos» розробила власне лінійне кодування ***SLN*** (***Sybyl Line Notation***) - універсальну мову представлення хімічної структури, яка дозволяє кодувати «стандартні» органічні сполуки, полімери і біомакромолекули. Основна різниця між SLN та SMILES – повна специфікація водневих атомів для усіх «неводневих» атомів у структурі.

The IUPAC International Chemical Identifier (InChI) – це формат опису молекул, який являється стандартом IUPAC (https://iupac.org/who-we-are/divisions/division-details/inchi). Базовий принцип, покладений в його основу – кожна речовина має одне і тільки одне опис. З огляду на те, що сучасні хімічні бази даних можуть містити описи мільйонів речовин, потрібні ефективні способи їх пошуку. InChI вирішує цю проблему, тому що являється комп'ютеризованим варіантом систематичної назви хімічної сполуки. InChI кодування було розроблено в рамках спеціального проекту ІЮПАК у 2000-2004 роках.

Так само, як і SMILES, опис молекули в InChI записується у вигляді текстового рядка, що дозволяє компактно описувати структурні формули. Наприклад, для етанолу InChI = 1S/C2H6O/c1-2-3/h3H,2H2,1H3.

До програмного забезпечення, яке підтримує представлення структури у вигляді нотації InChI і дозволяє генерувати InChI на основі структурної формули, відносяться:

1. редактори структурних формул:

а) *ChemDraw* (програмний пакет CambridgeSoft);
б) *Marvin Sketch* (програмний пакет ChemAxon);
в) *Symyx Draw* (програмний пакет Symyx Software);
г) *ChemSketch* (програмний пакет ACD Labs);

2. онлайн-редактори структурних формул:
   а) A new service for developers From the Royal Society of Chemistry https://developer.rsc.org;
   б) Chemical Identifier Resolver https://cactus.nci.nih.gov/chemical/structure;
   в) Online Sketcher пошукової системи *PubChem* https://pubchem.ncbi.nlm.nih.gov/edit3/index.html;
   г) Sketcher ресурсу ChemSpider http://www.chemspider.com; http://www.chemspider.com/StructureSearch.aspx.
   д) The IUPAC International Chemical Identifier (InChI) https://iupac.org/who-we-are/divisions/division-details/inchi.

Ресурс ChemSpider має засоби для роботи з кодами InChl та InChIKey, включаючи його пряму та зворотню конвертацію у формат MOLfile, перевірку валідності ідентифікатору, пошук інформації.

Код InChI має модульну структуру і складається з основних та внутрішніх шарів, які відокремлюються один від одного символом „/”. Головний шар містить брутто-формулу, інформацію про зв’язаність атомів, атоми водню.

Основні модулі (шари) InChI наступні: головний, ізотопи, стереохімія, таутомери. Така структура коду враховує необхідний об’єм інформації про сполуку, дозволяє використовувати його для проведення цілеспрямованого пошуку хімічної інформації. Єдиним обов’язковим елементом коду є формула. Нова інформація про сполуку включається у кінець коду у вигляді нового шару.

Код InChI, як і усі лінійні коди, являє собою рядок символів, але все одно він займає багато місця у базі даних. Тому для зручності зберігання інформації у комп’ютері використовується модифікація InChI – InChIKey, яка складається із двох частин. Останнім часом InChIKey використовується як аналог DOI для молекулярних структур.

Окремо слід згадати ***CAS registry number*** (він же CAS number, CAS RN, CAS #). Це унікальний чисельний ідентифікатор хімічних сполук, полімерів, біологічних послідовностей нуклеотидів або амінокислот, сумішей та сплавів, внесених до реєстру Chemical Abstracts Service. Номер CAS записується у вигляді трьох груп арабських чисел, розділених дефісами на три секції. Перша частина може містити до 7 цифр, друга містить дві цифри, третя складається з однієї цифри і виконує функцію контрольного символу. Номери призначаються в порядку зростання і не мають заздалегідь певного значення. Контрольна сума обчислюється шляхом додавання останньої цифри номера, помноженої на 1, другий праворуч цифри, помноженої на 2, третьої, помноженої на три і так далі до першої зліва цифри, завершуючи обчисленням залишку від ділення на 10. Наприклад, реєстраційний номер CAS для води 7732-18-5. Контрольна

сума обчислюється так: 8 × 1 + 1 × 2 + 2 × 3 + 3 × 4 + 7 × 5 + 7 × 6 = 105; 105 mod 10 = 5. Окремі ізомери молекул також отримують власний номер CAS. Наприклад, D-глюкоза має номер 50-99-7, L-глюкоза позначається 921-60-8; α-D-глюкоза – 26655-34-5 і т. і.

Хімічна реферативна служба (англ. Chemical Abstracts Service) – це підрозділ Американського хімічного товариства. Вона привласнює цей ідентифікатор всім речовинам, які коли-небудь згадувалися в літературі. Унікальний ідентифікатор призначений для більшої зручності пошуку хімічних сполук в літературі за рахунок усунення проблеми можливого різного найменування однієї сполуки. На даний час практично всі хімічні бази даних мають пошук за реєстраційним номером CAS.

Серед баз хімічних даних, які дозволяють використовувати лінійні коди як запит для пошуку інформації, можна вказати наступні:

1. NIST WebBook https://webbook.nist.gov/chemistry/inchi-ser;
2. NIH PubChem https://pubchem.ncbi.nlm.nih.gov;
3. ZINC15 UCSF http://zinc.docking.org/substances/home;
4. ChEBI https://www.ebi.ac.uk/chebi;
5. Compendium of Pesticide Common Names http://www.alanwood.net/pesticides;
6. ChemSpider http://www.chemspider.com;
7. AKos GmbH http://www.akosgmbh.de/AKosSamples;
8. Chemical Synthesis Database https://www.chemsynthesis.com;
9. EcoCyc https://biocyc.org/cpd-search.shtml.

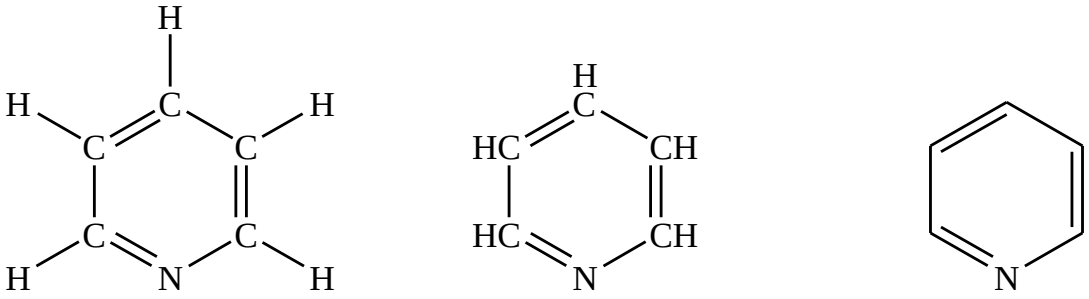
1.2.2 2-D Рівень представлення хімічних структур

Найбільш розповсюдженою міжнародною мовою для представлення хімічних структур є ***графічна – за допомогою структурних формул***. 2D структурна формула відображає атоми, наявні у молекулі, та зв’язки між ними (див. рис. 1.1).

Структурна формула – це різновид хімічної формули, який графічно описує розташування та порядок зв'язку атомів у сполуці і зображується на площині (рис. 1.2, тип 1). Зазвичай використовують структурні формули, де зв'язки з атомами водню не позначаються валентними рисками (рис. 1.2, тип 2). В іншому типі структурних формул (скелетних), які застосовуються для великих молекул в органічній хімії, атоми водню, зв'язані з атомами карбону, не вказуються, також не позначаються атоми карбону (рис. 1.2, тип 3).

Для вирішення конкретних завдань були розроблені різні методи представлення хімічних структур. Наприклад, в патентах для позначення і захисту загальної структури цілого ряду (серії) хімічних сполук використовуються ***структури*** (формули) ***Маркуша*** (родові хімічні структури). Формула Маркуша не відповідає конкретній сполуці, вона є зручним способом позначення хімічних структур в узагальненому вигляді. На практиці їй може відповідати

велика кількість конкретних сполук. Формула Маркуша включає один або кілька варіантів замісників, що поєднуються в групу альтернативних структур.



Тип 1 Тип 2 Тип 3

Рисунок 1.2 – Типи структурних формул піридину

Графічна формула – це зручне і зрозуміле хіміку представлення хімічної сполуки. Однак структурні формули є незручними з точки зору зберігання, пошуку і аналізу комп'ютерними методами.

З математичного погляду, структурні формули можна розглядати як топологічний граф, побудований із вузлів або вершин (атомів), з'єднаних ребрами або дугами (хімічні зв'язки). Отже, багато проблем, що виникають при обробці структурної інформації, можна віднести до сфери математичної теорії графів. У теорії графів широко використовуються різні матриці, такі як матриці суміжності чи зв'язаності, і їх також можна застосувати для представлення інформації про хімічну структуру. Проте, вони представляють лише зв'язки всередині структури і зазвичай не містять інформацію про типи атомів і порядки зв'язків. На основі матриць було побудовано таблицю зв'язаності, яку було запропоновано у 1965, і яка домінувала у 80-х роках минулого століття як форма представлення хімічної структурної інформації в комп'ютерних системах. Стандартний формат обміну даними на основі таблиці зв'язаності є ***Molfile формат***, розроблений ***MDL Information Systems***, який використовується у хімічних графічних редакторах.

На даний час наявні пакети програм, можливості яких дозволяють будувати структурні формули хімічних сполук і перетворювати їх у форму, зручну для подальшого комп'ютерного зберігання.

Як правило, кожний графічний редактор хімічних формул має свій власний формат організації збереження структурної інформації на 2D-рівні. Однак для обміну даними між популярними програмами використовується стандартний формат *.mol. На рисунку 1.3 наведено типову структуру *.mol файлу на прикладі молекули етанолу.

Ключовими елементами типового *.mol файлу є

- інформація про програму і дату створення файлу,
- інформація про кількість атомів і кількість зв'язків;
- декартові координати на площині для кожного атому,
- тип кожного атому;
- таблиця зв'язків.

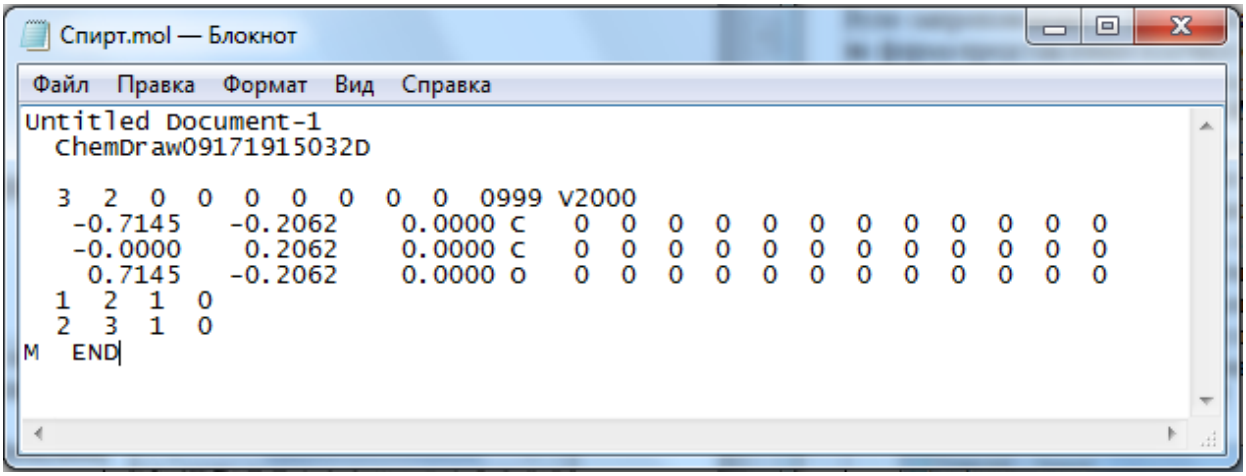


*Рисунок 1.3 – Типова структура *.mol файлу*

Графічний редактор хімічних формул Marvin Sketch є складовою частиною програмного пакету Marvin фірми ChemAxon (http://www.chemaxon.com). Типова структура робочого вікна програми MarvinScketch представлена на рисунку 1.4. Вікно містить основні елементи: рядок заголовку, рядок меню, рядок стану, смуги прокрутки, робочу область, вертикальні і горизонтальні панелі інструментів. За бажанням можна змінити зовнішній вигляд робочого вікна шляхом використання команди ***Configurations*** меню ***View*** програми.

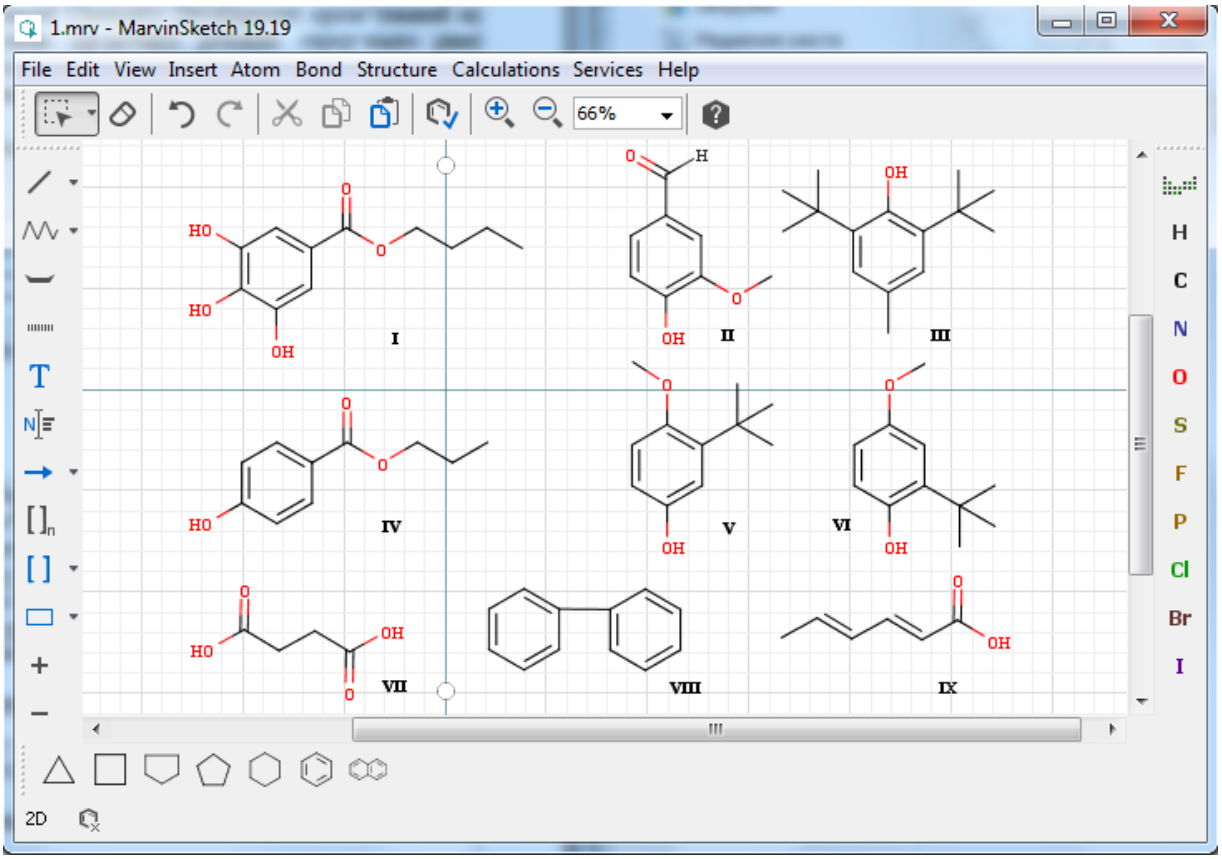


Рисунок 1.4 – Структура робочого вікна програми Marvin Sketch

Основні можливості програми:

– створення структурних формул органічних речовин;

– генерування назви сполуки;

– створення схем реакцій;

– перетворення 2D-структурного ескізу в 3D-модель;

– вільне обертання в площині і 3D-обертання моделей;

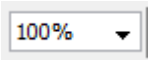
– розрахунок молекулярної маси та елементного складу структури або структурного фрагмента;

– генерування і візуалізація структурної інформації у різноманітних файлових форматах: MOL, MOL2, SDF, RXN, RDF (V2000/V3000), SMILES, SMARTS/SMIRKS (recursive), MRV, InChI, CML, PDB;

– експорт створених файлів у формати PNG, JPEG, BMP, PNF.

Основні інструменти програми MarvinScketch наведені в таблиці 1.1.

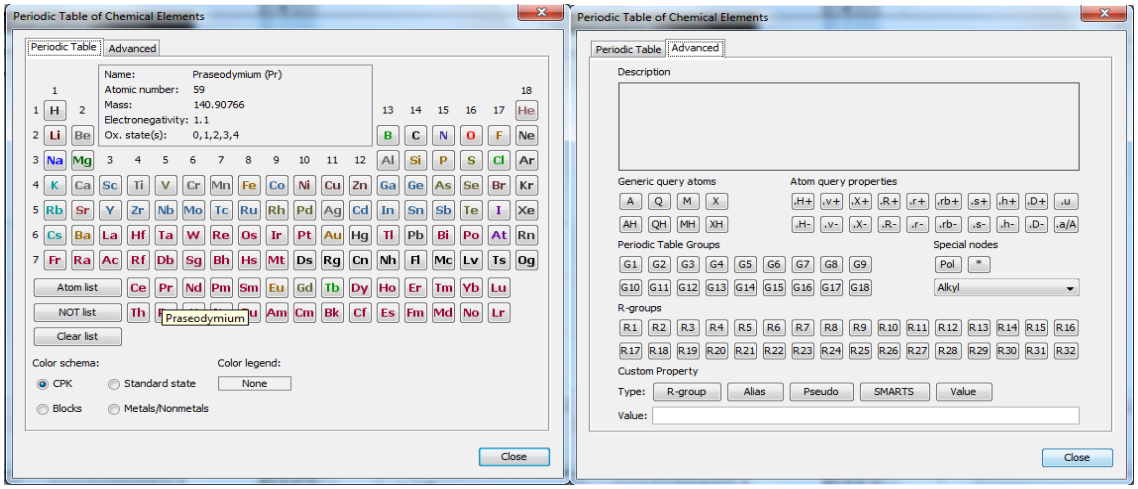
Таблиця 1.1 – Основні інструменти програми MarvinScketch

Кнопка	Назва	Кнопка	Назва
	Ласо		Зв’язок
	Ластик		Ланцюг
	Відміна / Повтор		Стереозв’язок
	Ножиці		Пунктирний зв’язок
	Копія		Текст
	Вставка		Редактор метки
	Перевірка структури		Стрілка
	Збільшення/зменшення		Дужки
	Масштаб		Графіка

Права вертикальна панель робочого вікна програми містить основні найпоширеніші елементи, які можна додавати до структури. Також можна викликати інтерактивну періодичну систему елементів (рис. 1.5, а) та вибрати необхідний елемент, якщо він відсутній на основній панелі. В програмі є вкладка, що містить шаблони груп та спеціальних символів (рис. 1.5, б).

Інтерфейс програми є нескладним і досить зрозумілим, тому не потребує багато часу для освоєння. Панелі інструментів програми забезпечують швидкий доступ до основних команд побудови зв’язків, готових шаблонів, елементів хімічної графіки.

Доступ до повного переліку шаблонів програми MarvinScketch здійснюється через меню ***Insert*** команда ***Template Library*** (рис. 1.6).



а *б*

Рисунок 1.5 – Інтерактивна періодична система елементів (а) та шаблони груп і спеціальних символів (б) програми MarvinSketch

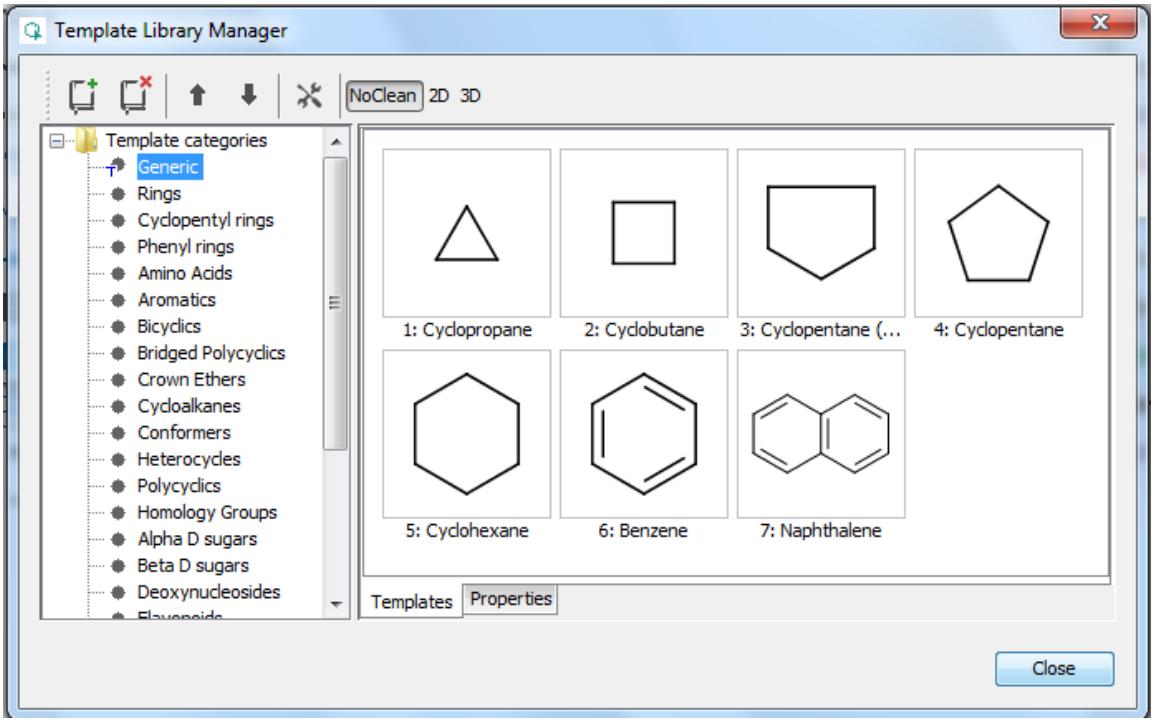


Рисунок 1.6 – Бібліотека шаблонів (Template Library) програми Marvin Sketch

Меню ***Insert*** програми також містить доступ до основних інструментів побудови зв'язків усіх типів, ланцюгів, стрілок, дужок та інших елементів хімічної графіки.

Команда ***Groups*** меню ***Insert*** програми дає можливість додати до структури замісник простої будови, тоді як для побудови складних фрагментів молекули може бути використана команда ***New Substituent***, яка відкриває одночасно нове вікно програми, де і будується окрема складний структурний фрагмент.

В програмі MarvinSketch меню ***Calculations*** програми дозволяє для виділеної структури

– розрахувати молярну масу і склад (команда ***Elemental Analysis),***

– згенерувати назву за номенклатурою ІЮПАК і традиційну назву (команда ***Naming***),

– прогнозувати величину pK_a (команда ***Protonation),***

– прогнозувати параметри ЯМР ^1H і ^{13}C спектрів (команда ***NMR***),

– прогнозувати кількість і структуру ізомерів (команда ***Isomers)*** і конформерів (команда ***Conformaton).***

Лінійний код хімічної структури можна згенерувати за допомогою команди ***Copy As Smiles*** меню ***Edit***.

Графічний редактор хімічних формул ChemDraw є одним із компонентів інтегрованого пакету програм ***ChemOffice*** (http://www.cambridgesoft.com/). Він призначений для створення структурних формул хімічних сполук будь-якої складності, схем хімічних реакцій і елементів хімічної графіки. Це одна з найбільш поширених програм, що використовуються для відображення структурних формул хімічних сполук.

Робоче вікно програми містить наступні елементи: рядок заголовку, рядок основного меню, вертикальну і горизонтальну смуги прокрутки, панелі інструментів, робочу область і рядок стану (рис. 1.7).

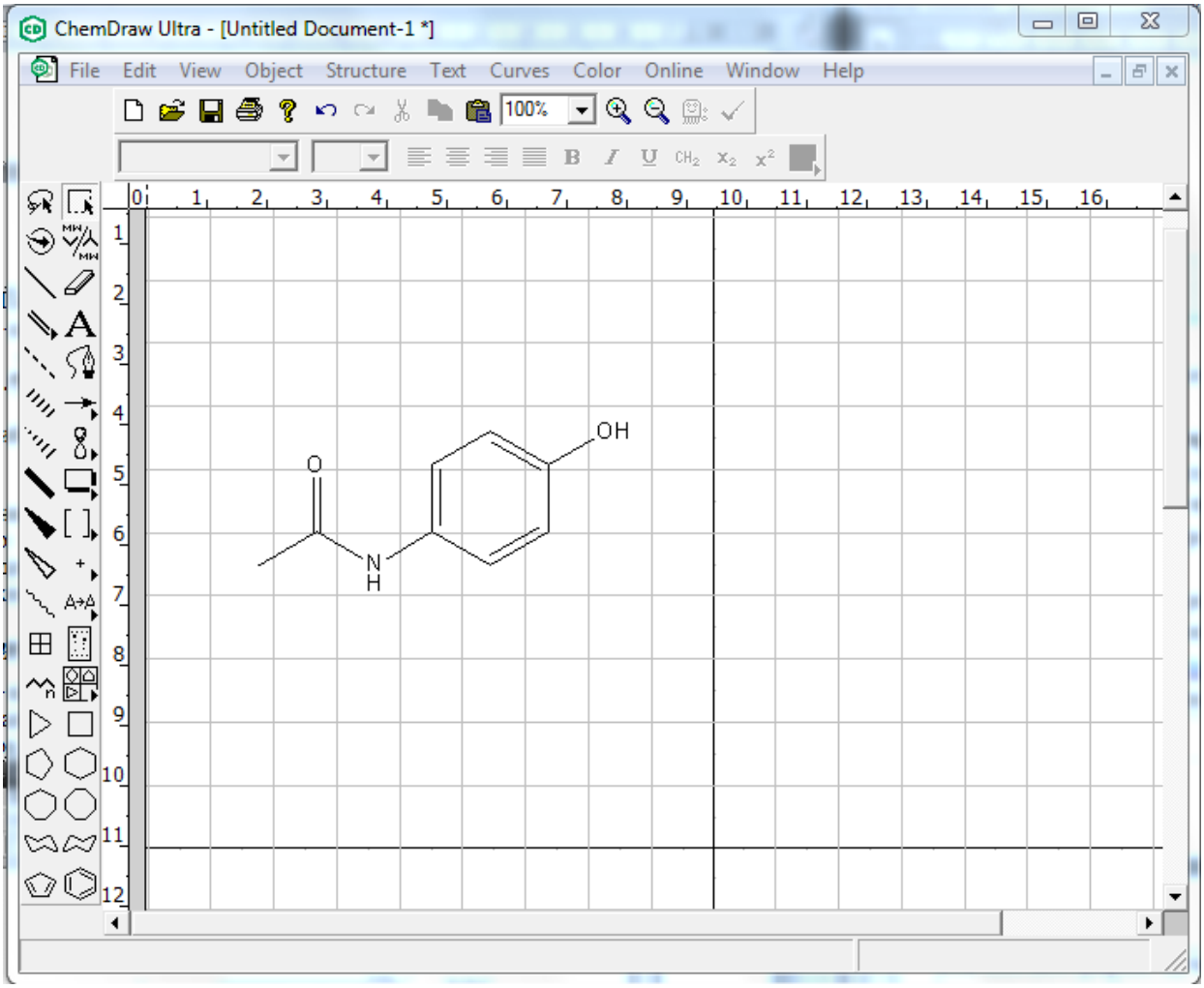


Рисунок 1.7 – Структура робочого вікна хімічного редактора ChemDraw

Інтерфейс програми інтуїтивно зрозумілий, тому не потребує забагато зусиль до освоєння. Хімічні інструменти розташовані на лівій вертикальній панелі. До основних інструментів можна віднести ***Ласо, Прямокутник, Зв'язок, Ластик, Множинний зв'язок, Текст, Пунктирний зв'язок, Перо, Реакційна стрілка, Стереозв'язок, Орбіталі*** і т. і.

Програма дає можливість працювати з файлами різних форматів: CDX, SKC, JDX, MOL. Містить широкий спектр шаблонів хімічних структур і елементів хімічної графіки. Для хімічних структур можливі розрахунок і прогнозування великої кількості параметрів.

В програмі ChemDraw для генерування назви сполуки на основі виділеної структурної формули використовується команда ***Convert Structure to Name*** в меню програми ***Structure***. Команда ***Convert Name to Structure*** здійснює генерування структурної формули на основі назви сполуки за номенклатурою ІЮПАК.

Для виділеної структури можна прогнозувати параметри ЯМР спектрів за допомогою відповідних команд ***1H-NMR Shifts*** і ***13C-NMR Shifts*** в меню програми ***Structure***.

Команда ***Copy as...*** меню ***Edit*** дозволяє згенерувати і розмістити в буфер обміну інформацію про хімічну сполуку у вигляді лінійних кодів SMILES та SLN.

Через меню ***View*** програми (команда ***Show Analysis Window***) можна викликати окреме вікно для розрахунку деяких параметрів на основі структурної формули: молекулярної маси, елементного складу тощо.

Графічний редактор хімічних формул ChemSketch з пакету програм ACD/Labs канадської фірми Advanced Chemistry Development орієнтований на роботу зі структурними формулами органічних речовин середнього рівня складності. Він має велику бібліотеку готових формул, але в ньому зручно складати також хімічні формули неорганічних речовин.

Основні можливості програми:

– створення структурних нескладних формул органічних речовин;

– створення структур речовин на основі лінійних кодів InChI і SMILES;

– створення структур Маркуша;

– створення схем реакцій;

– перетворення 2D-структурного ескізу в 3D-модель;

– проведення пошуку хімічних структур у файлах різного формату через комп'ютерну файлову систему;

– проведення повного або часткового структурного пошуку;

– розрахунок молекулярної маси та елементного складу структури або структурного фрагмента, оцінка макроскопічних властивостей – молекулярної рефракції, молярного об'єму, густини та ін.;

– експорт створених файлів у різні формати, зокрема, PDF та HTML.

Структура робочого вікна програми зображена на рисунку 1.8.

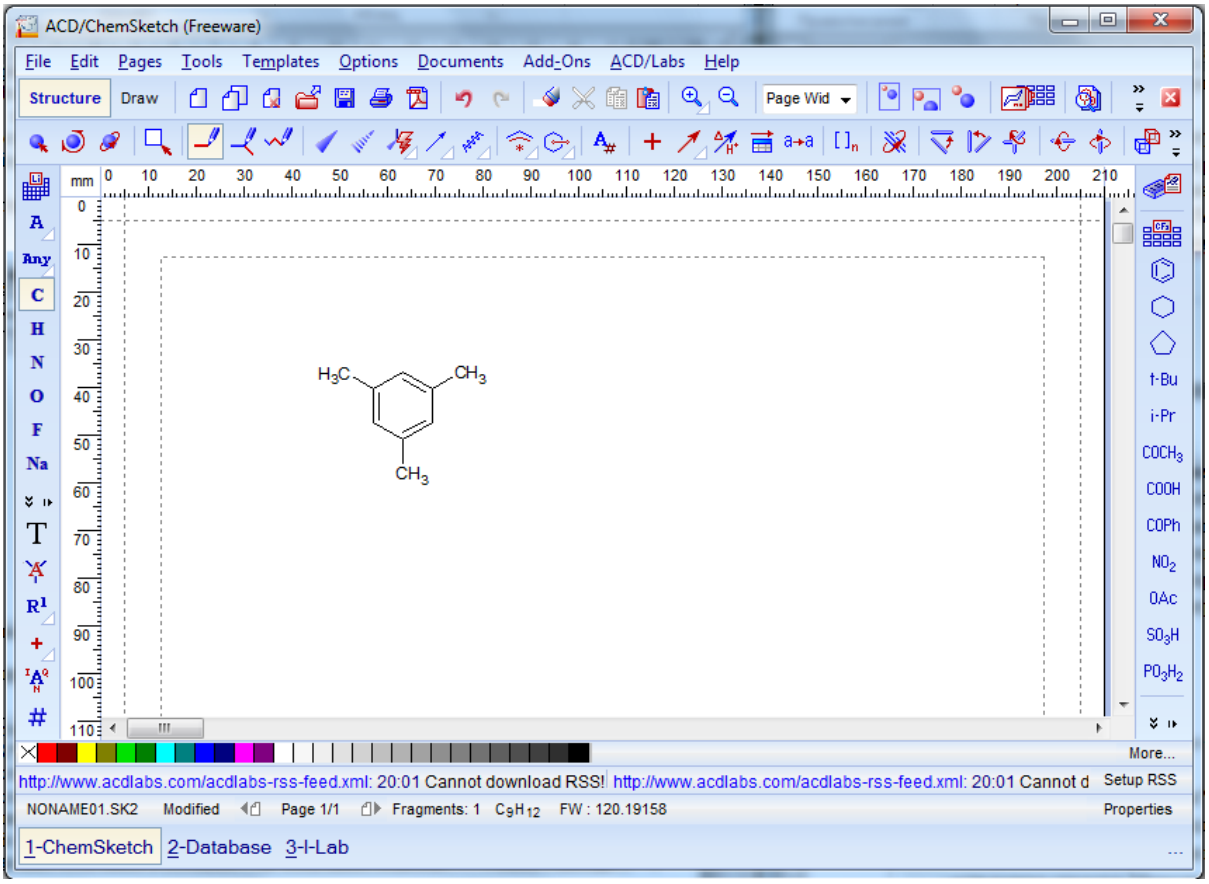


Рисунок 1.8 – Структура робочого вікна хімічного редактора ChemSketch

Панель інструментів редактора розташована нижче рядка меню програми і містить кнопки швидкого доступу до команд ***Збереження*** і ***Відкриття Файлів, Відміну*** і ***Повтору Дії, Видалення, Вирізання*** і ***Копіювання Фрагмента, Масштабування***, а також до ряду специфічних команд. На цій панелі знаходяться також кнопки ***Structure*** і ***Draw***, які перемикають режим роботи редактора.

Редактор ChemSketch працює в двох режимах: ***Structure*** (Структура) і ***Draw*** (Малювання), що відрізняються призначенням і набором інструментів. У режимі ***Структура*** створюються структурні формули речовин, схеми і рівняння реакцій. У режимі ***Структура*** над робочою областю знаходиться панель інструментів, яка містить кнопки команд для створення і роботи з хімічними формулами.

Режим ***Малювання*** зручний для введення і редагування текстових блоків, створення таблиць, схем, нестандартних графічних об'єктів. У режимі ***Малювання*** активною є панель інструментів для роботи з графічними об'єктами.

Панель ***Атоми*** розташована зліва від робочої області і активна лише в режимі ***Структура***. Вона включає інструмент доступу до таблиці елементів Д.І. Менделєєва, кнопки, що позначають атоми, й інструменти для встановлення властивостей атома (заряд, валентність, номер тощо).

У режимі ***Малювання*** замість цієї панелі активною є панель інструмен-

тів ***Автофігури***.

Панель ***Радикали*** розташована праворуч від робочої області. Вона містить таблицю радикалів, що дозволяє використовувати готові шаблони радикалів – різних структурних фрагментів. Активна ця панель лише в режимі ***Структура***.

Робоча область програми являє собою окрему сторінку. Документ ChemSketch може містити одну або декілька таких сторінок. Їх у свою чергу можна копіювати і видаляти, використовуючи команди меню ***Pages*** програми.

Рядок стану містить інформацію про загальне число сторінок, номер робочої сторінки, і кнопки для навігації між сторінками. Також включає довідкову інформацію: назву поточного файлу, число хімічних структур у робочій області, молекулярну формулу і масу виділеного фрагмента структури.

Формули сполук створюються в режимі ***Структура***. Основою формули органічної сполуки є вуглеводневий скелет. Для його створення редактор містить три інструменти: ***Draw Normal*** (Малювати нормально), ***Draw Continuous*** (Малювати безперервно) і ***Draw Chains*** (Малювати ланцюжки). Щоб малювати саме вуглеводневі структури, окрім вибору інструмента треба вибрати на панелі атомів символ атома вуглецю. Будь-який з цих інструментів малювання зв’язків при кликанні на лінії зв'язку перетворить одинарний зв'язок у подвійний, потім у потрійний і знову в одинарний.

Циклічні структури можна малювати інструментом ***Draw Continuous.*** Щоб замкнути цикл, останній клік здійснюють на першому атомі вуглецю. Для циклів з великим числом атомів вуглецю можна використовувати інший спосіб: за допомогою інструмента ***Draw Chains*** малюють ланцюжок з потрібним числом атомів вуглецю, потім, використовуючи ***Draw Normal***, з’єднують перший атом з останнім. У результаті отримують неоптимізовані цикли, що складаються з одних ліній. Щоб представити їх у правильному вигляді, використовують команду ***Clean Structure*** (F9) в пункті ***Tools*** головного меню програми. Для перетворення формули Кекуле в ароматичну структуру використовується команда меню ***Tools*** – ***Show Aromaticity*** (показати ароматичність), зворотна операція виконується командою ***Hide Aromaticity.***

Для додавання нескладних циклічних структур можна скористатися відповідним шаблоном у таблиці радикалів (права панель).

Інструмент для редагування і вставки групи знаходиться на панелі ***Атоми*** – команда ***Edit Atom Label***. Для додавання групи атомів потрібно активізувати цю кнопку кліканням по атому в структурній формулі, який має бути замінений на групу або по вільному місцю сторінки поряд з формулою. При цьому відкривається діалогове вікно ***Edit Label*** (рис. 1.9). У цьому вікні необхідно вибрати в наявному списку формулу потрібної групи атомів або її умовне позначення, при цьому вибрана формула з'являється у вікні редагування. Якщо потрібної формули немає, її можна ввести у вікні редагування, при цьому індекси вводяться в рядок (спеціальна кнопка дозволяє вводити букви грецького алфавіту). Наприкінці треба натиснути кнопку ***Insert*** (Вставити).

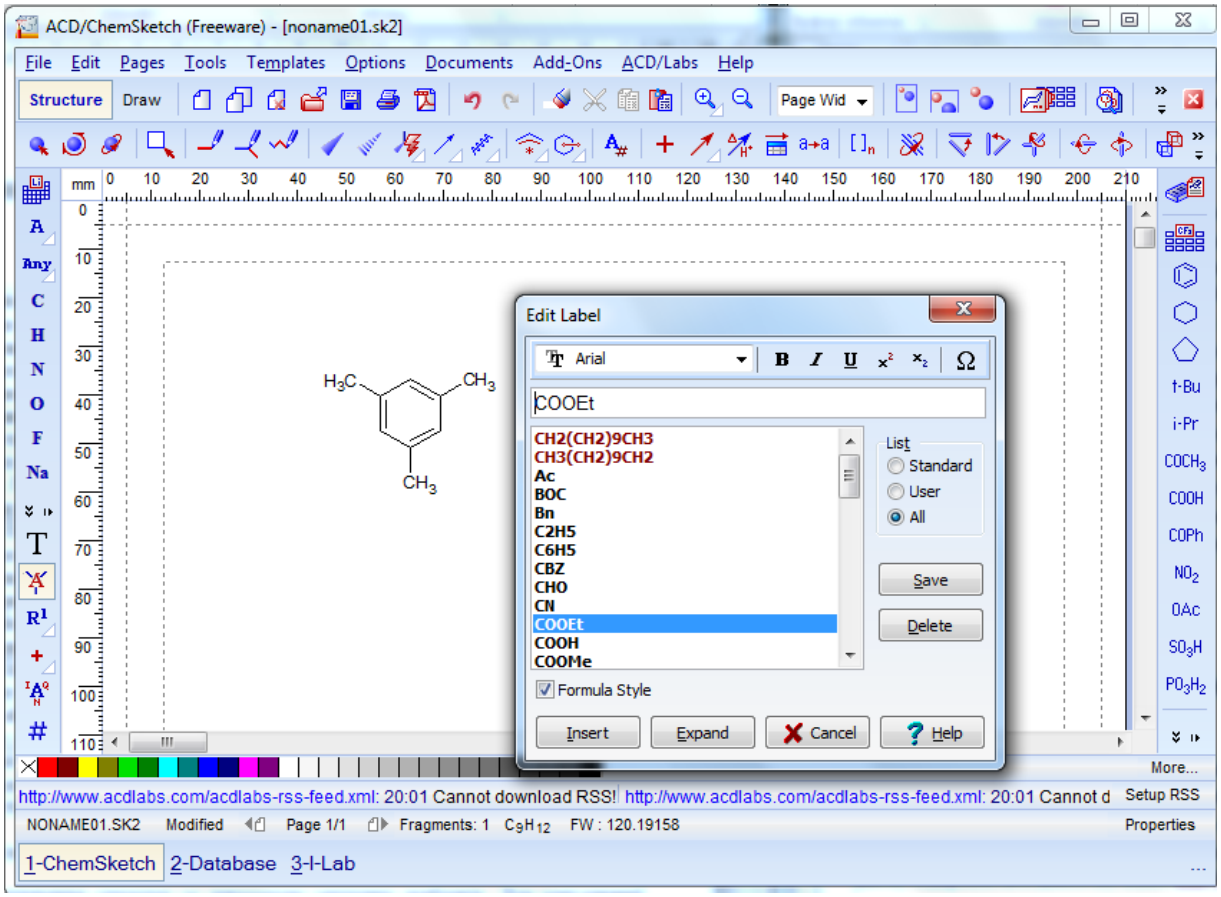


Рисунок 1.9 – Діалогове вікно Edit Label хімічного редактора ChemSketch

Готові структурні формули можна обертати, змінювати положення окремих атомів або фрагментів, змінювати розміри структури в цілому. Для переміщення формули або структурного фрагмента використовується інструмент ***Select/Move*** (Вибір/Переміщення). Після активації інструмента необхідно перетягуванням змінити положення атома або зв'язка, або цілої структури на робочій області.

Для копіювання структурних формул та інших виділених об'єктів використовуються звичайні прийоми (команда ***Copy*** меню ***Edit*** програми, одночасне використання клавіш ***Ctrl+C,*** кнопки панелі інструментів, перетягування при натиснутій клавіші ***Ctrl).*** При виконанні процедури вставки об'єкта (команда ***Paste*** меню ***Edit*** програми, поєднання клавіш ***Ctrl+V,*** кнопки панелі інструментів) до курсору прикріплюється шаблон структури. Наступний клік у робочій області призводить до створення копії даної структури.

Також можна використовувати миттєві шаблони. Для цього необхідно активізувати на панелі інструментів кнопку ***Instant Template*** (Миттєвий шаблон) і клікнути по відповідному елементу готової структури. Цей елемент прикріплюється до курсору. Наступний клік по робочій області додає такий шаблон. Для відміни шаблонування використовують клік правою кнопкою миші або кнопку ***Esc*** клавіатури.

Інструмент ***Select/Size/Rotation*** (Вибір/Розмір/Обертання) забезпечує обертання структури у площині. Якщо виділити рамкою структуру або її фрагмент, то з'являються маркери зміни розміру фрагмента.

Інструмент ***3D-Rotation*** (3D-Обертання) дозволяє обертати структуру або її фрагмент у просторі.

Можна змінювати положення структури або певного фрагмента відносно обраного зв'язку. Для цього використовують інструменти ***Set Bond Vertically*** (встановити зв'язок вертикально), ***Set Bond Horizontally*** (встановити зв'язок горизонтально) і ***Flip on Bond*** (відобразити відносно зв'язка). Кнопки ***Flip*** ... (відображення зверху вниз та зліва направо) потребують попереднього виділення структури чи фрагмента. В іншому випадку відбудеться відображення усієї сторінки.

Кнопка ***Change Position*** дозволяє змінити положення символів атома водню відносно вуглецевого атома.

Для швидкого переходу до стандартного вигляду структури її необхідно виділити і скористатися кнопкою ***Clean Structure***, яка дозволяє вирівняти кути і довжини зв'язків у межах площини. Використання інструмента ***3D Optimisation*** приведе до повного розгортання формули.

Відображення усіх атомів водню, зв'язаних з вуглецевим скелетом структури, встановлюється за допомогою команди ***Add Explicit Hydrogens*** меню ***Tools***. Команда ***Remove Explicit Hydrogens*** відповідного меню відміняє цю дію.

В програмі ChemSketch для генерування назви сполуки на основі виділеної структурної формули використовується команда ***Name for Structure*** меню ***Tools/Generate***.

Меню ***Tools/Generate*** також дозволяю згенерувати лінійні коди SMILES і InChI.

Характеристики хімічної сполуки (брутто-формулу, молярну масу, елементний склад) можна розрахувати за допомогою меню ***Tools/Calculate.***

Окрім одинарних, подвійних і потрійних зв'язків у формулах іноді потрібні інші типи – координаційні, водневі та ін.

Інструменти ***Up Stereo Bonds*** і ***Down Stereo Bonds*** дозволяють перетворити звичайну структуру у її просторове зображення. Послідовно клікаючи по стереозображенню зв'язка можна змінити його вигляд, однак у цьому режимі не відображується кратність зв'язків – одинарні, подвійні і потрійні мають однаковий вид. Для повернення в звичайний режим будування використовується кнопка ***Draw Normal***.

Донорно-акцепторні зв'язки, водневі і координаційні зв'язки зображуються відповідно стрілками або пунктирними лініями. Для їх будування використовується інструмент ***Coordinating Bonds***. Він містить додаткову палітру з різними типами ліній та стрілок. Напрямок стрілки можна змінити послідовним кліканням на зв'язок. Для повернення у звичайний режим необхідно клікнути правою кнопкою миші на вільному місці робочого листа, або вибрати кнопку ***Draw Normal.***

Аналогічно працюють з інструментом ***Special Bonds,*** який дозволяє об-

рати ще один з декількох можливих способів представлення хімічного зв'язку. Вид кнопки буде залежати від попередньо обраного інструмента.

Створення складних хімічних формул навіть у спеціалізованому хімічному редакторі потребує значних витрат часу. Для зменшення цих витрат ChemSketch містить велику кількість вбудованих готових шаблонів для найскладніших структурних формул, а також інших рисунків хімічного змісту. Для відкриття вікна вибору шаблонів використовується команда ***Template Window*** меню ***Templates*** (Шаблони) програми, клавіша ***F5,*** або відповідна кнопка на панелі інструментів (рис. 1.10).

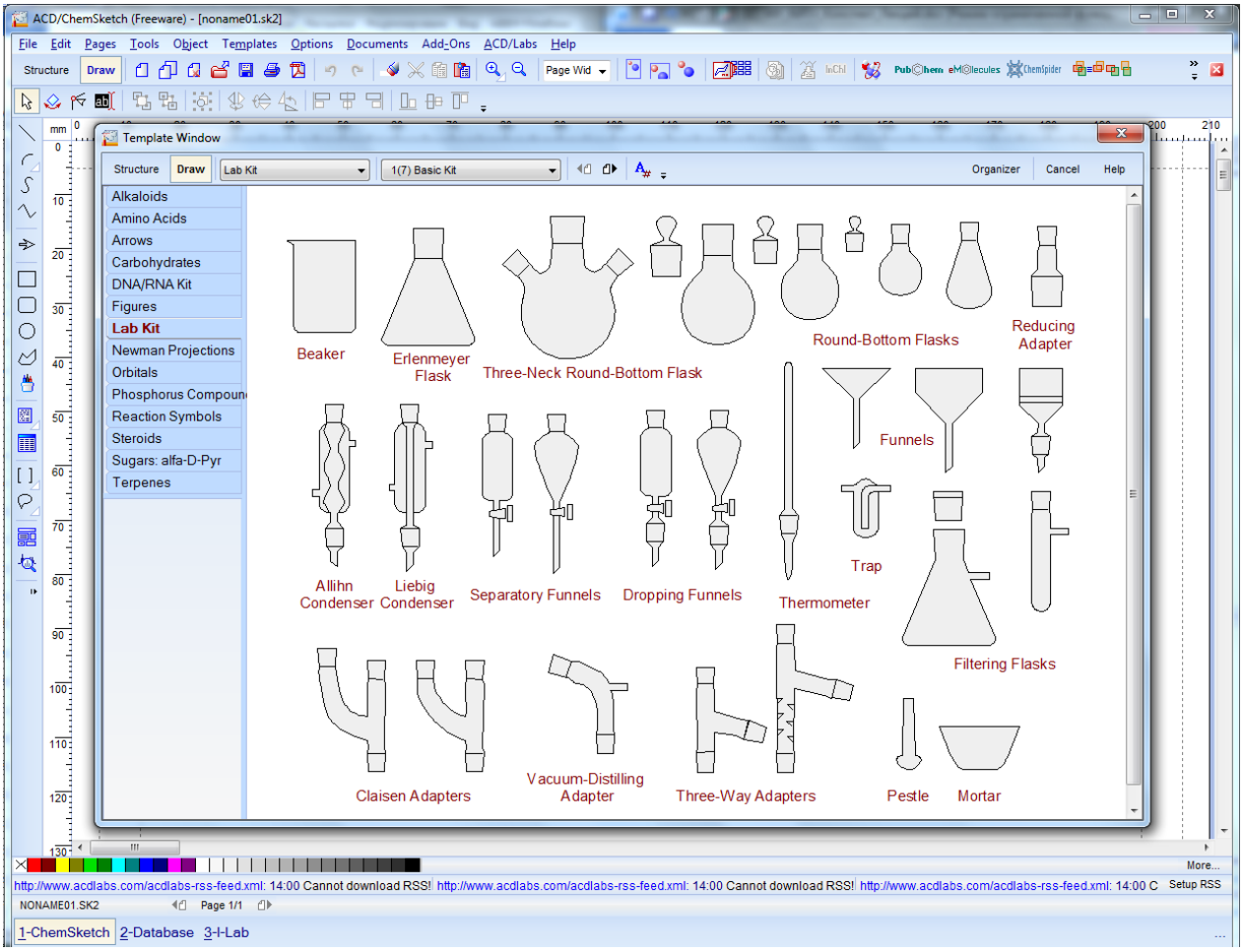


Рисунок 1.10 – Діалогове вікно Template Window хімічного редактора ChemSketch

1.2.3 3-D Рівень представлення хімічних структур

Всі молекули є тривимірними об'єктами, тому досить часто виникають ситуації, коли треба враховувати явище ***стереоізомерії***. Це можливе при використанні ***3D-моделей*** хімічних структур.

3D-Рівень представлення хімічної структури – представлення хімічних

атомів і молекул на рівні їх просторових моделей.

Молекулярна модель – модель, що використовується при експериментальному та теоретичному вивченні молекул, при дослідженні їх електронної структури та взаємодій. Модель може бути розрахунковою або може відображати справжній фізичний об'єкт.

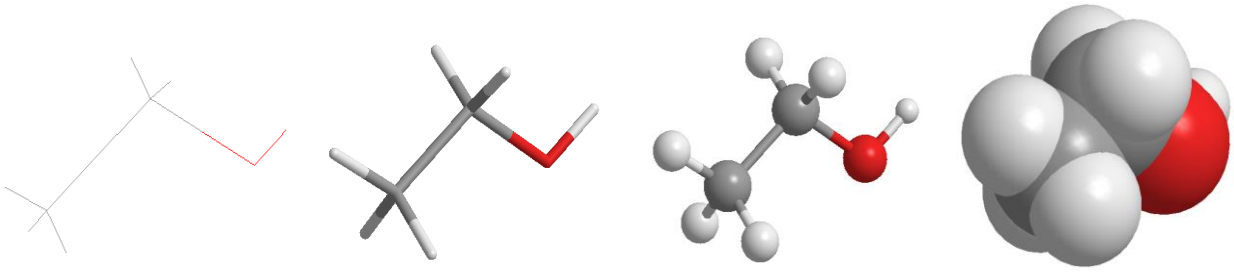
За способом представлення молекулярні моделі можна поділити на наступні: каркасні, стрижньові, кульково-стрижньові, об'ємні, стрічкові та багатогранники.

Каркасні моделі показують зв'язки між атомами молекули у вигляді тонких ліній (у вигляді каркасу) (рис. 1.11, *а*).

Стрижньові моделі показують об'ємні зв'язки між атомами молекули. Один із варіантів каркасних моделей (рис. 1.11, *б*).

Кульково-стрижньові моделі представляють атоми та зв'язки між ними (рис. 1.11, *в*).

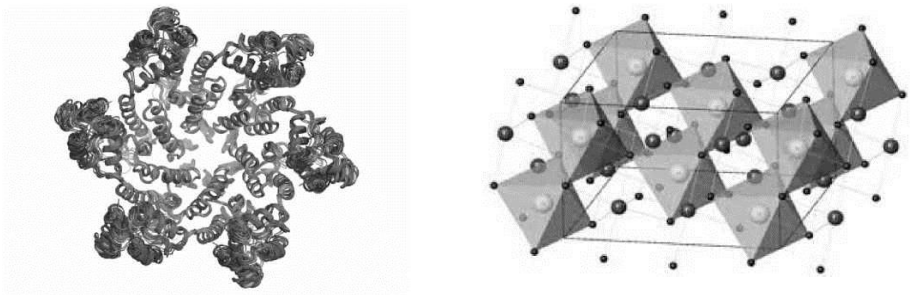
Об'ємні моделі відтворюють відносні атомні розміри в молекулі та форму молекули (рис. 1.11, г).



а *б* *в* *г*

а – каркасна модель; б – стрижньова модель;
в – кульково-стрижньова модель; г – об'ємна модель
Рисунок 1.11 – 3D-молекулярні моделі етанолу

Існують також ***стрічкові моделі***, які відтворюють просторову будову складних біомолекул (рис. 1.12, а) та ***багатогранники***, які використовуються для представлення моделей кристалічної будови речовини (рис. 1.12, б).



а *б*

а – стрічкова модель; б – багатогранники
Рисунок 1.12 – Типи 3D-молекулярних моделей

На 3D-рівні представлення структурної інформації оперують такими поняттями як молекулярна конфігурація, рівноважна молекулярна конфігурація, конформація, конформер.

Молекулярна конфігурація – це просторове розміщення атомів молекули згідно величин структурних параметрів.

Рівноважна молекулярна конфігурація – це конфігурація молекули, що відповідає мінімуму її повної енергії.

Конформація – одне із неідентичних розташувань у просторі груп атомів даної молекули, що утворюється внаслідок обертання навколо одинарного зв'язку без його розриву зі збереженням стереохімічної конфігурації молекули.

Конформери – конформаційні ізомери, яким відповідають мінімуми на поверхні потенціальної енергії, і які знаходяться у рівноважному стані, хоча не можуть бути виділені через невисокий енергетичний бар'єр обертання (8-40 кДж/моль).

3D-Структуру можна представити у вигляді ***декартових*** або ***внутрішніх координат***.

Одним із поширених способів представлення 3D-структури є її представлення у Декартовій системі координат (***Cartesian Coordinates***). При цьому наводиться інформація про точні ***x, y*** та ***z*** координати кожного атома молекули. Таким чином будується таблиця, в якій кожному атому відповідає окремий рядок, а кожній координаті – окремий стовпчик (рис. 1.13).

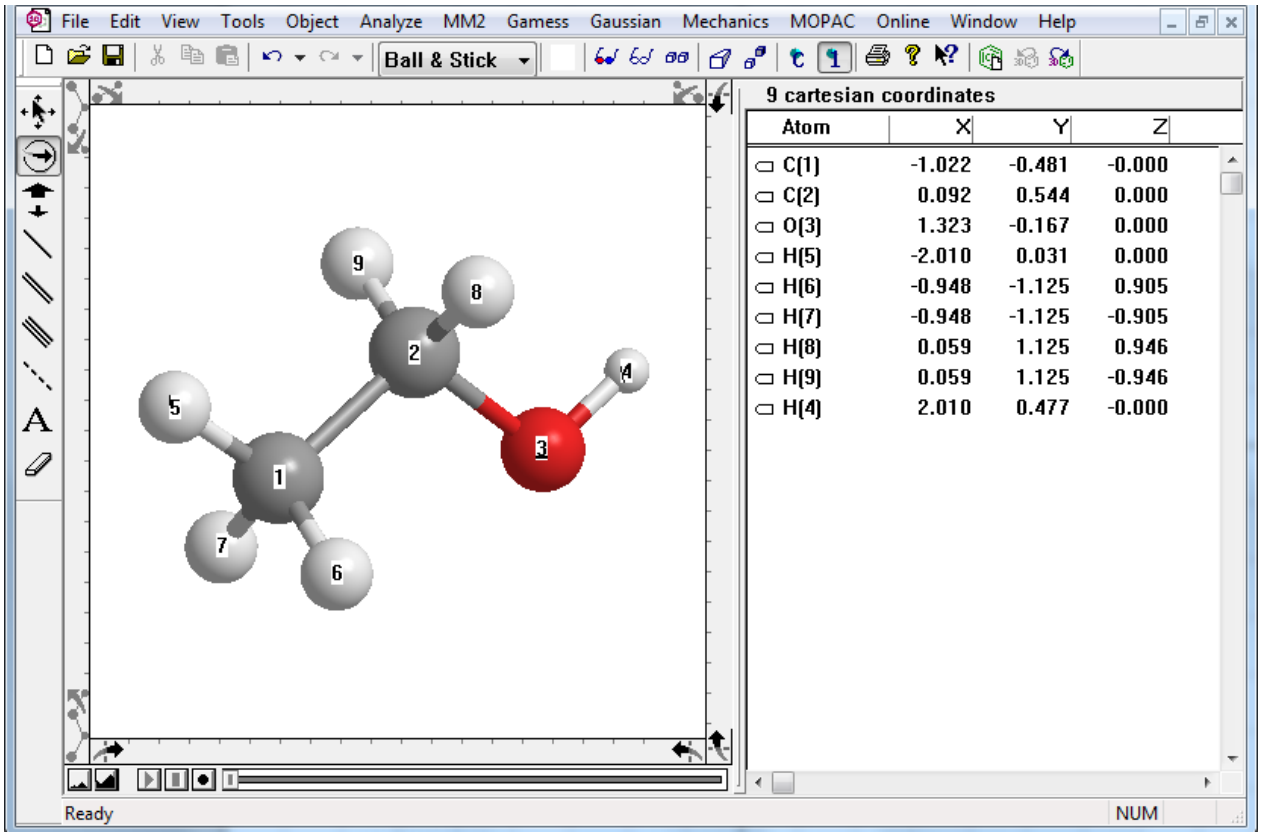


Рисунок 1.13 – Декартові координати молекули етанолу

Інформація про 3D-структуру молекули у вигляді Декартових координат є складовою файлів форматів *.mol та *.xyz.

Другим найпоширенішим методом представлення 3D-структури молекули є використання ***внутрішніх координат*** – довжин зв'язків, величин валентних і торсійних кутів. Внутрішні координати представляють просторове розміщення атомів хімічної частинки відносно один одного. Це представлення називають ***Z-матриця*** (рис. 1.14).

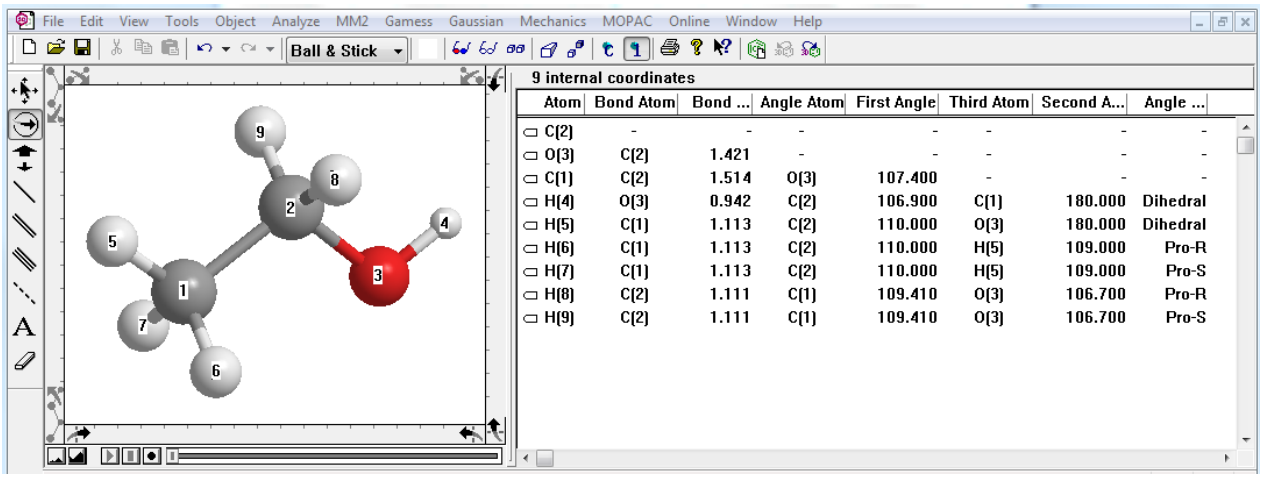


Рисунок 1.14 – Внутрішні координати (Z-матриця) молекули етанолу

Для правильної побудови Z-матриці а також її інтерпретації необхідно дотримуватися декількох правил. Кожний рядок Z-матриці описує один атом молекули. Перший стовбець матриці містить перелік усіх атомів даної молекули, другий – величини довжин валентних зв'язків, четвертий – величини валентних кутів, шостий – величини торсійних кутів. Перший атом (C2), інформація про який наведена у першому рядку, розташовується на початку системи координат. Другий атом (O3) знаходиться від нього на відстані 1,421 Å (див. рис. 1.14). Третій атом (C1) знаходиться на відстані 1,514 Å від першого атому C2. Величина валентного кута між атомами 1–2–3 (C1–C2–O3) складає 107,400 °. Торсійний кут, який наведено у останньому стовпчику значень Z-матриці, – це кут, що утворюється між площинами відповідних атомів. На рисунку 1.14 перший торсійний кут 180,000 ° утворюється між площинами, які містять атоми H4–O3–C2 і O3–C2–C1, відповідно.

За виключенням перших трьох атомів, кожний атом у Z-матриці представлений набором із трьох внутрішніх координат. Загальна кількість параметрів, наведених у Z-матриці, може бути розрахована як 3N-6, де N – кількість атомів у складі молекули.

Z-Матриця найчастіше використовується для введення вихідної структурної інформації у квантово-хімічних розрахунках, оскільки вона зручно представляює просторову будову молекули (молекулярну конфігурацію).

У вигляді Z-матриці інформація про 3D-структуру хімічної частинки є складовою файлів форматів *.zmt, *.gjf, *.dat і *.mop.

При дослідження структури молекул важливе значення має вибір початкової молекулярної геометрії, так як це значною мірою визначає якість подальших досліджень. Основними методами генерування просторових координат молекулярних структур є:

– використання даних кристалографічних баз рентгеноструктурних даних;

– підбір стандартної геометрії із бібліотеки фрагментів;

– перетворення структур з двовимірної у тривимірну модель шляхом автоматизованих процедур.

Найбільш якісні результати можна отримати при ***використанні даних рентгеноструктурного аналізу*** (РСА). Найважливішою базою кристалографічних даних є Кембріджська кристалографічна база структурних даних (The Cambridge Crystallographic Data Centre, CCDC, https://www.ccdc.cam.ac.uk/solutions/csd-system/components/csd). База містить експериментально визначені атомні координати для органічних і неорганічних сполук, постійно оновлюється і містить дані про більш ніж 1000000 сполук (рис. 1.15). Кембріджський центр пропонує платні послуги і програмне забезпечення для пошуку інформації у своїх ресурсах і аналізу результатів. Результатом такого пошуку є файл, що містить інформацію про просторову будови відповідної молекули.

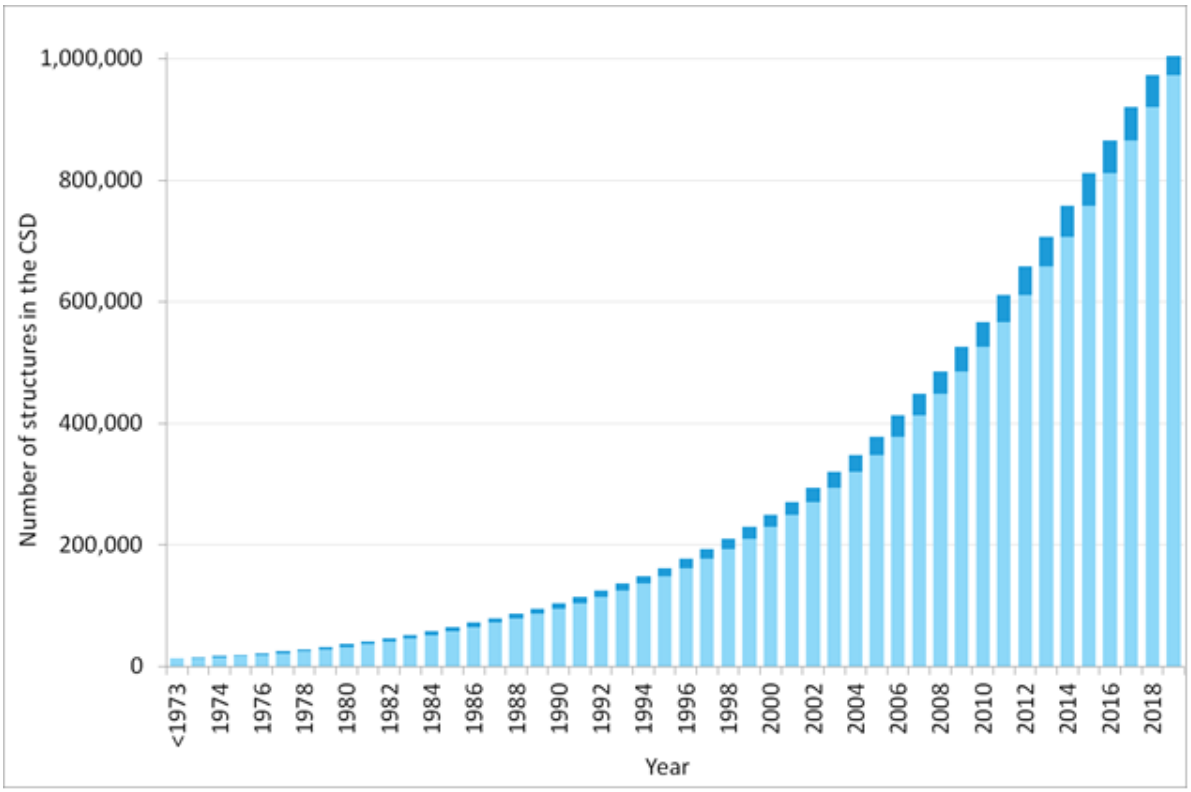


Рисунок 1.15 – Динаміка росту кількості даних РСА в базі CCDC

Для полімерних біомолекул існує ряд спеціалізованих банків структурної інформації: Банк даних білкових молекул (Protein Data Bank, PDB, https://www.wwpdb.or), База даних нуклеїнових кислот (Nucleic Acid Database,

NDB, http://ndbserver.rutgers.edu).

Головним джерелом даних про будову неорганічних сполук є База структурних даних неорганічних кристалів (Inorganic Crystal Structure Database, ICSD, https://www.fiz-karlsruhe.de/en/produkte-und-dienstleistungen/inorganic-crystal-structure-database-icsd).

Другий метод, який часто використовується для побудови молекулярних структур, базується на використанні молекулярних фрагментів, попередньо зібраних у відповідні бібліотеки. Цей метод використовують, коли немає доступу до кристалографічної бази даних, або кристалографічна інформація відсутня. На даний час переважна більшість програм для молекулярного моделювання містять такі бібліотеки і надають можливість генерувати 3D структури молекул.

Перетворення структур з двовимірної у тривимірну шляхом автоматизованих процедур також дає можливість одержати просторову модель молекули. Для цього можна використовувати наступні програмні продукти: MarvinSpace, 3D Viewer, Chem3D.

3D структуру хімічної частинки можна згенерувати засобами програмного комплексу ACD/Labs (компонент ***3D Viewer),*** побудувавши 2D структурну формулу за допомогою інструментів ChemSketch. Аналогічно працюють програмні комплекси Chem Axon (2D структура будується компонентом MarvinSketch, а 3D генерується продуктом ***MarvinSpase)*** і ChemOffice (2D структура будується у ChemDraw, а 3D-модель генерується компонентом ***Chem3D).***

MarvinSpace – програма для швидкої візуалізації просторової молекулярної конфігурації хімічних сполук. Вона входить до складу програмного комплексу Chem Axon (http://www.chemaxon.com) і повністю інтегрована з відповідним графічним редактором хімічних формул MarvinScketch.

Програма Marvin Spase здатна до візуалізації біомолекул, структурна інформація про які міститься у форматі *pdb. Саме цей формат є основним для зберігання кристалографічної інформації в структурних базах даних, таких як CSDB, Protein Data Bank.

3D Viewer – програма для швидкої візуалізації просторової молекулярної конфігурації хімічних сполук. Вона входить до складу програмного комплексу ACD/Labs і повністю інтегрована з відповідним графічним редактором хімічних формул ChemSketch. Це дозволяє генерувати 3D-моделі хімічних сполук на базі їх 2D-структурних формул. Серед основних можливостей програми 3D Viewer слід зазначити наступні:

– можливість керування 3D-моделями: переміщення, обертання, зміна розмірів, кольорів, стилів;

– відображення молекули у вигляді просторових моделей різних типів: каркасної, стрижневої, кульково-стрижневої, у вигляді сфер або дисків;

– візуалізація поверхонь Ван-дер-Ваальса;

– вимірювання та редагування геометричних параметрів об’єкта: довжин зв’язків, валентних кутів, торсійних кутів;

– оптимізація молекулярної геометрії методом молекулярної механіки;

– автономне обертання об’єкта в просторі;
– експорт 3D-моделей в інші програмні продукти.

Робоче вікно програми 3D Viewer (рис. 1.17) містить:

– рядок меню;
– панель інструментів;
– робочу область;
– рядок стану, який містить інформацію про ім’я поточного файлу та брутто-формулу активної структури;
– екранні кнопки, які дозволяють швидкий перехід та обмін даними між вже завантаженими програмами комплексу Acd/Labs.

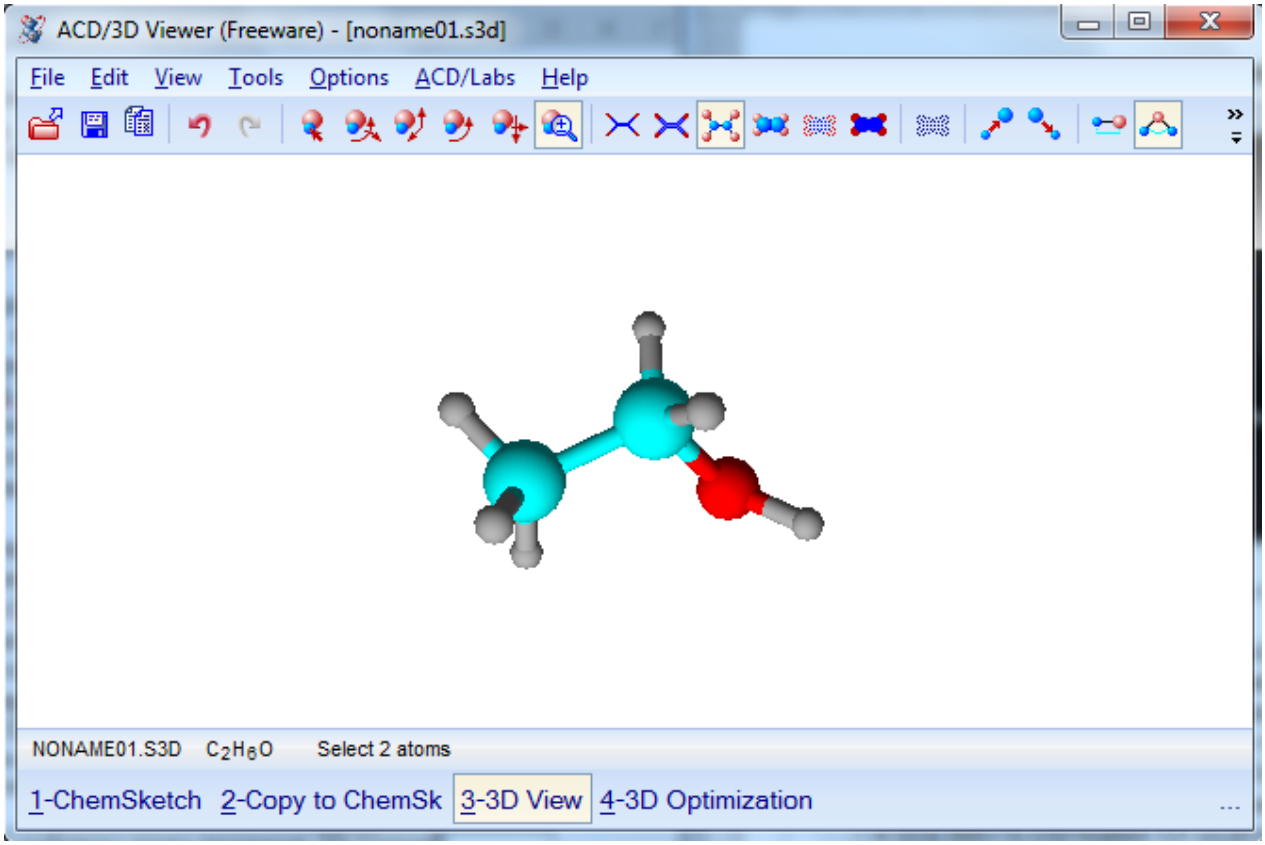


Рисунок 1.17 – Робоче вікно програми 3D Viewer

Визначення відстані між окремими атомами структури, торсіонних кутів і кутів між зв'язками здійснюється в спеціальному режимі 3D перегляду – в програмі 3D Viewer. Для перемикання в даний режим слід в нижньому меню робочого вікна пакету ACD/Labs вибрати відповідний пункт. При цьому молекулярна структура, що знаходиться на робочому полі, буде скопійована в вікно перегляду (див. рис. 1.17).

Панель інструментів в режимі перегляду представлена кількома блоками кнопок, функції яких дублюються в підпунктах основного меню. Перший блок містить кнопки: ***Відкрити файл***, ***Зберегти файл***.

У другому блоці знаходяться шість кнопок, які регламентують режими переміщення/обертання структури.

Далі розташовані кнопки, що дозволяють перемикати режими відображення молекулярної структури у вікні перегляду: ***каркасна модель, стрижньова модель, кульково-стрижньова модель, об'ємна модель, точкова модель, дискова модель***. Слідом за ними розташована кнопка вмикання/вимикання відображення прозорої об'ємної моделі, потім – кнопка ***збільшення і зменшення радіусів атомів*** (для деяких варіантів відображення структури).

За допомогою кнопок наступного блоку здійснюється:

- вимірювання відстані між атомами (рис. 1.18);
- вимірювання кутів між зв'язками (рис. 1.19);
- вимірювання торсійних кутів.

Якщо молекулярна структура побудована не за даними РСА, то перед виконанням вимірювань необхідно обов'язково оптимізувати просторову орієнтацію молекулярної структури. Це здійснюється за допомогою команди ***3D оптимізації*** в меню ***Tools***.

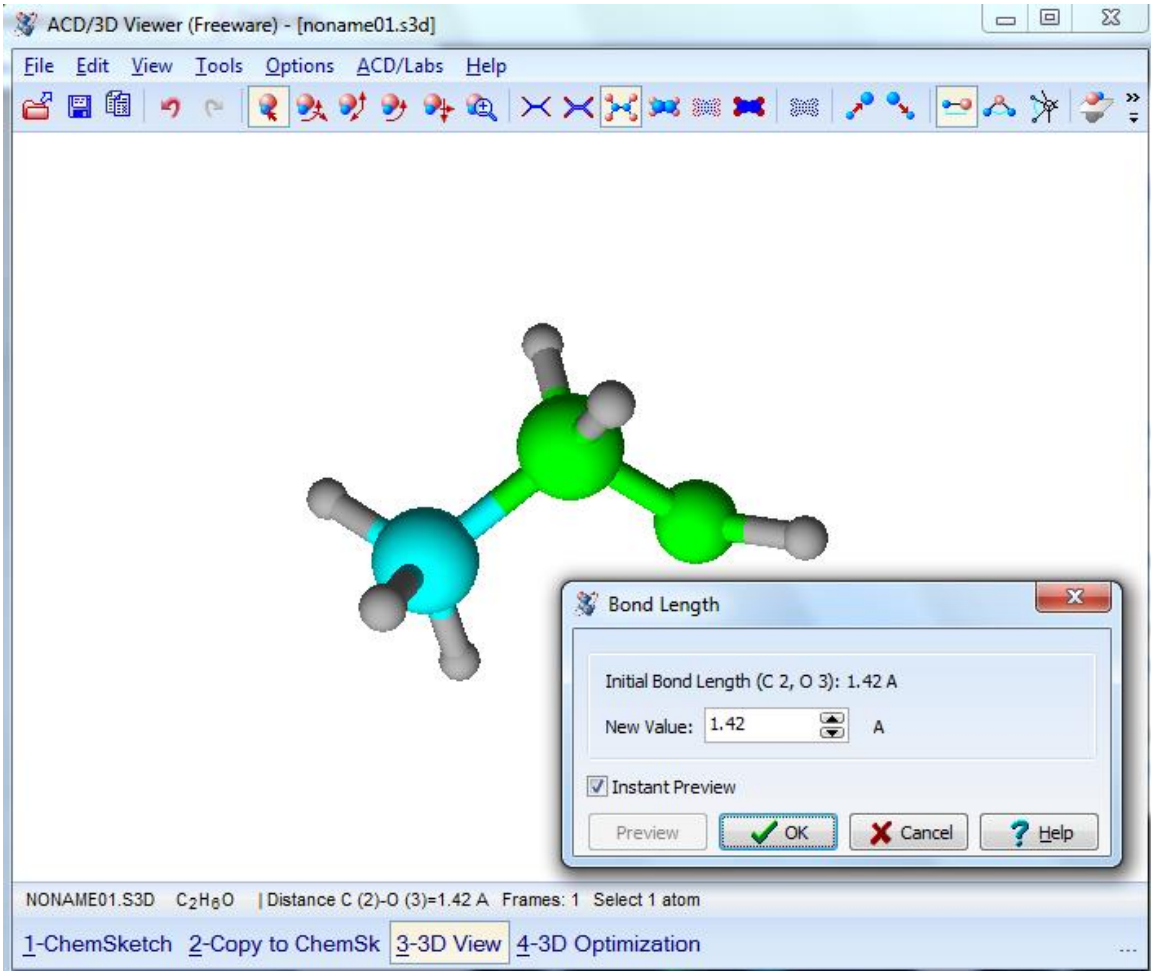


Рисунок 1.18 – Вимірювання довжини зв'язку

Для вимірювання відстані між атомами (довжин зв'язків) необхідно натисканням активувати відповідну кнопку ***Bond Length*** і послідовно кліками виділити два атоми структури – вони при цьому змінюють колір на зелений

(див. рис. 1.18). В області повідомлень відобразиться інформація про відстань між виділеними атомами.

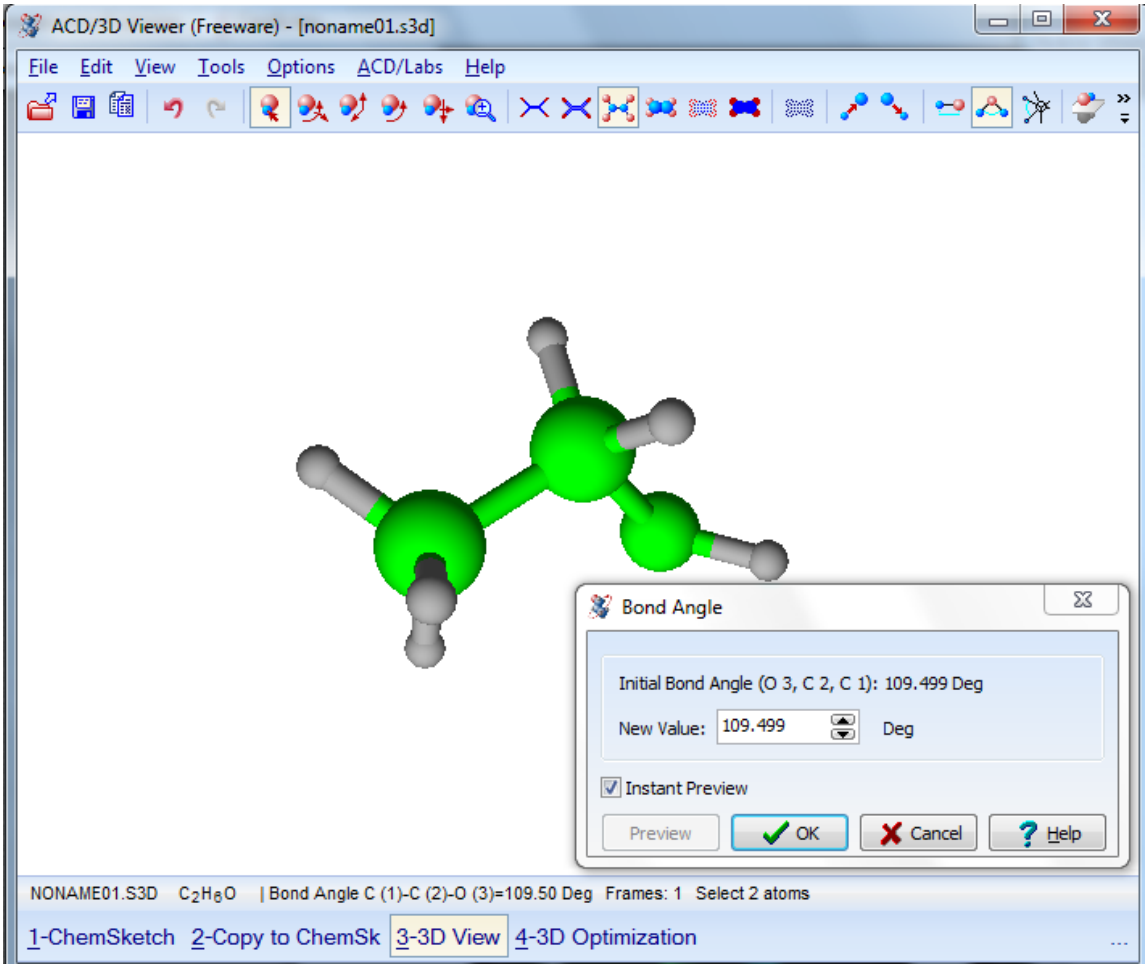


Рисунок 1.19 – Вимірювання валентного кута

Вимірювання кутів між зв'язками здійснюється за допомогою відповідної кнопки ***Bond Angle*** на панелі інструментів з послідуючим виділенням трьох атомів (див. рис. 1.19).

При вимірюванні торсіонних кутів необхідно активувати кнопку ***Torsion Angle*** на панелі інструментів і виділити чотири атоми (рис. 1.20).

Для зручності роботи в програмі передбачені можливості ***зміни кольорів*** атомів і фону вікна перегляду (кнопка ***Set Colors*** на панелі інструментів), ***автообертання молекули*** (крайня права кнопка ***Auto Rotate*** на панелі інструментів) і ***автообертання молекули зі зміною стилю структури*** (меню ***Tools*** команда ***Auto Rotate and Change Style***).

Для зручності роботи передбачені можливості ***зміни кольорів*** атомів і фону вікна перегляду (кнопка ***Set Colors*** на панелі інструментів), ***автообертання молекули*** (крайня права кнопка ***Auto Rotate*** на панелі інструментів).

Програма 3D Viewer дозволяє створювати анімації з обертанням молекулярних структур і зберігати їх у форматі GIF.

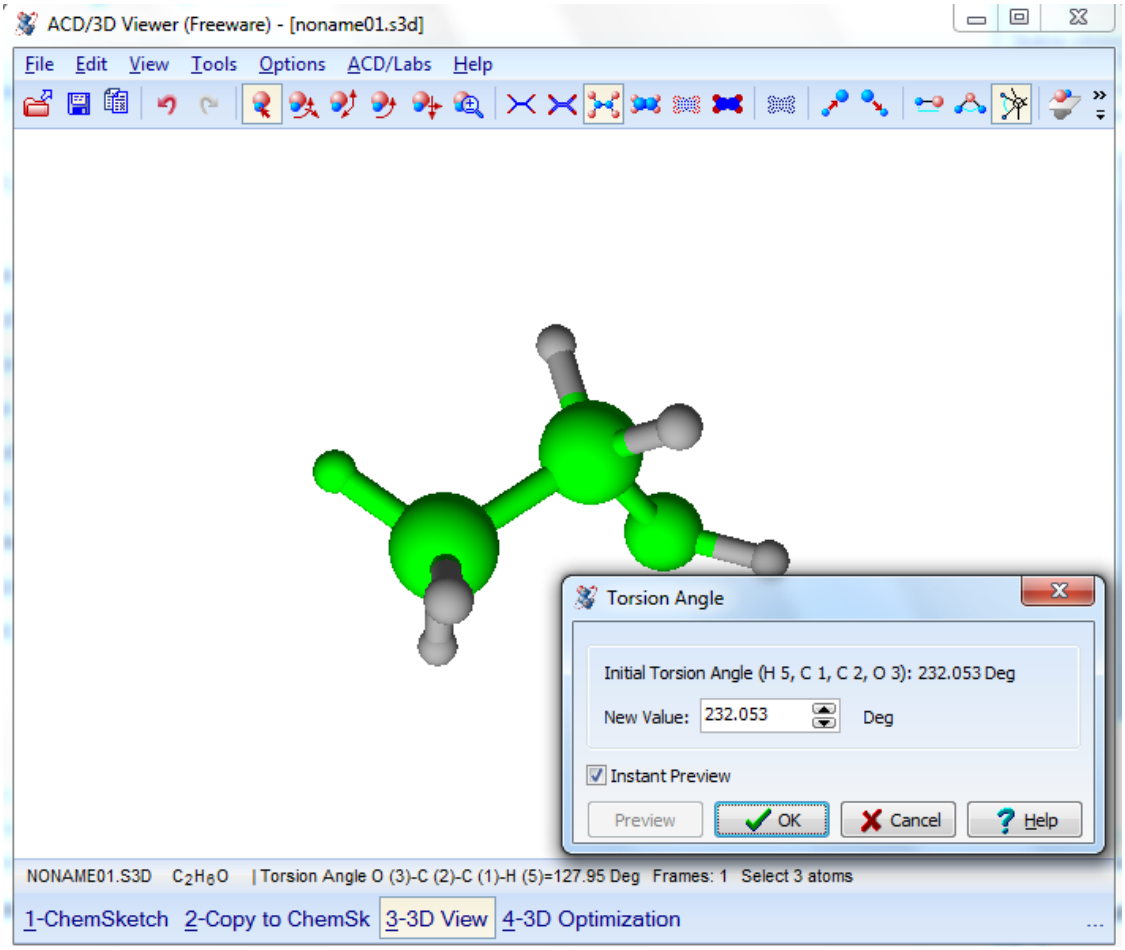


Рисунок 1.20 – Вимірювання торсійного кута

1.3 Інтеграція даних спеціалізованих програм до програм пакету MS Office

Microsoft Office – офісний пакет програм, створених корпорацією Microsoft для операційних систем Microsoft Windows, Windows Phone, Android, macOS, iOS. До складу цього пакету входить програмне забезпечення для роботи з різними типами документів: текстами, електронними таблицями, базами даних тощо. Microsoft Office поставляється в різних редакціях, які відрізняються складом пакету і ціною.

Найпоширеніші програми, що входять до пакету, – це текстовий процесор Microsoft Word, табличний процесор Microsoft Excel, персональний комунікатор Microsoft Outlook, програма підготовки презентацій Microsoft PowerPoint, програма для управління базами даних (СУБД) Microsoft Access.

Програма **Word** призначена для роботи з різноманітними текстовими документами. Word дозволяє:

– створювати текстові документи будь-якої складності та спрямування: наукові; юридичні; фінансово-економічні; художні тощо;

– редагувати та форматувати документи з метою полегшення їх розуміння і читання;

– додавати до текстів різноманітні ілюстративні матеріали: таблиці, діаграми, фотографії, малюнки та схеми тощо;

– використовувати численні та різноманітні вбудовані інструменти (для створення векторних рисунків, таблиць, діаграм, інструменти для редагування растрових малюнків, тощо), що дозволяють виконувати багато завдань без додаткового залучення спеціалізованих програмних продуктів;

– зберігати документи не тільки в форматі Word, але й в інших форматах, наприклад, в форматі PDF, HTML або шаблонів, що суттєво полегшує публікацію та розповсюдження документів;

– використовувати вбудовану мову програмування Visual Basic for Applications (VBA) для створення повнофункціональних додатків з інтерфейсом користувача у вигляді екранних форм з елементами керування і багато ін.

Але ця програма незручна для створення хімічних формул, особливо органічних сполук. При підготовці докладів, документів, звітів хімічного напрямку більш раціонально хімічні формули створювати у спеціальних хімічних редакторах, а потів вбудовувати до відповідного документу в форматі Word.

В документи, створені програмою Word, можуть бути додані різноманітні об'єкти, що дає можливість зробити документи більш змістовними і зрозумілими.

Об'єкти, створені програмами MarvinSketch, ChemDraw та ChemSketch, можна додати до документу Word методом копіювання. Спочатку цей об’єкт треба виділити на робочому полі відповідної програми і активувати команду ***Copy*** (***Копіювати***) у меню програми або натиснути ***Ctrl+C***. Потім треба підвести курсор миші до потрібного місця в документі Word і активувати команду ***Past*** (***Вставити***) або натиснути ***Ctrl+V***.

При копіюванні об’єкту із робочого вікна програми MarvinSketch при використанні команди ***Copy As*** (***Копіювати Як***) є можливість вибрати формат, в якому буде представлена інформація в документі Word (рис. 1.21). Але слід враховувати, що графічне зображення структурної формули хімічної сполуки в цільовому документі Word можна отримати тільки при використанні форматів ***Image*** (див. рис. 1.21). Об’єкт при цьому вбудовується в документ як рисунок і не може в подальшому редагуватися інструментами програми MarvinSketch.

При копіюванні хімічної формули або іншого об’єкту із робочого вікна програми ChemDraw об’єкт в документ Word вбудовується як об’єкт CS ChemDraw Drawing (рис. 1.22) з можливістю подальшого редагування інструментами програми ChemDraw. Редагування можна здійснювати після активації команди ***Edit*** (***Редагування***) або ***Open*** (***Відкрити***) (див. рис. 1.22)

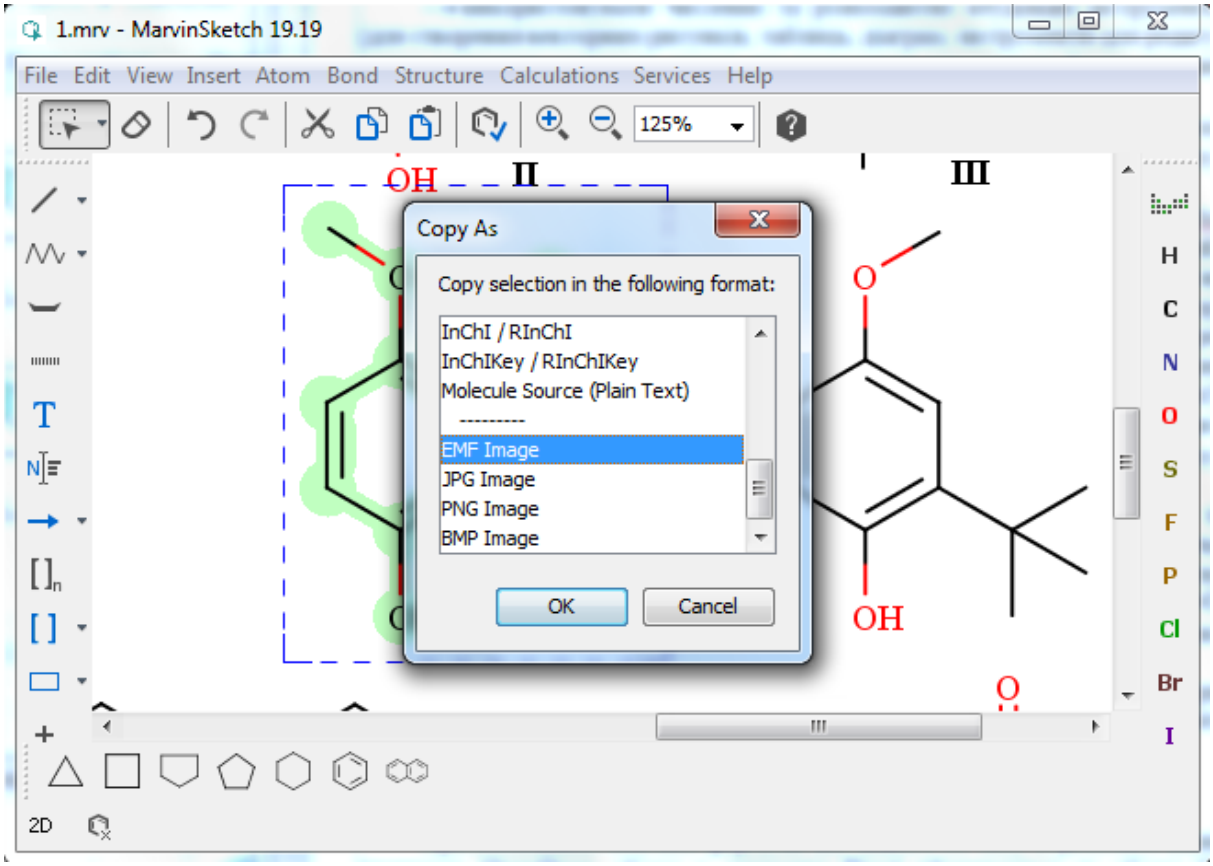


Рисунок 1.21 – Копіювання об’єктів із програми MarvinSketch

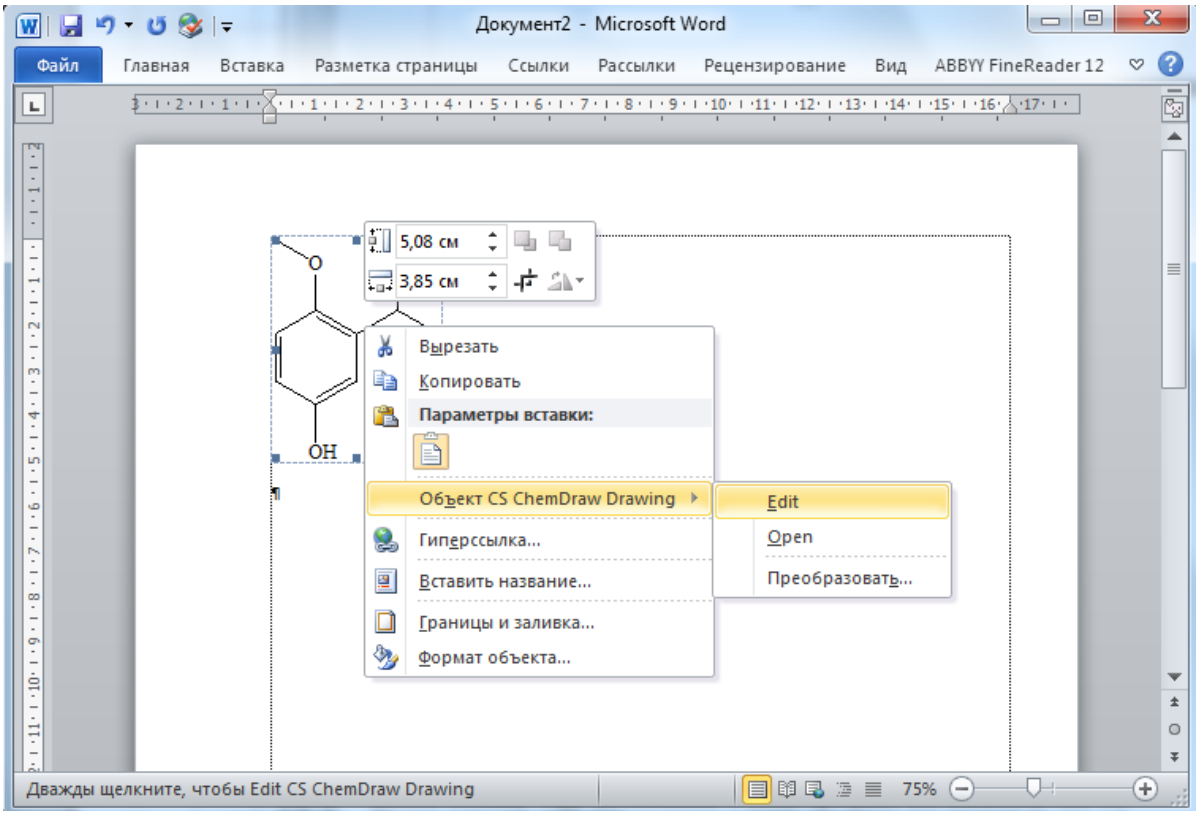


Рисунок 1.22 – Копіювання об’єктів із програми ChemDraw

При копіюванні хімічної формули або іншого об’єкту із робочого вікна програми ChemSketch об’єкт в документ Word вбудовується як об’єкт ChemSketch з можливістю подальшого редагування інструментами програми ChemSketch (рис. 1.23).

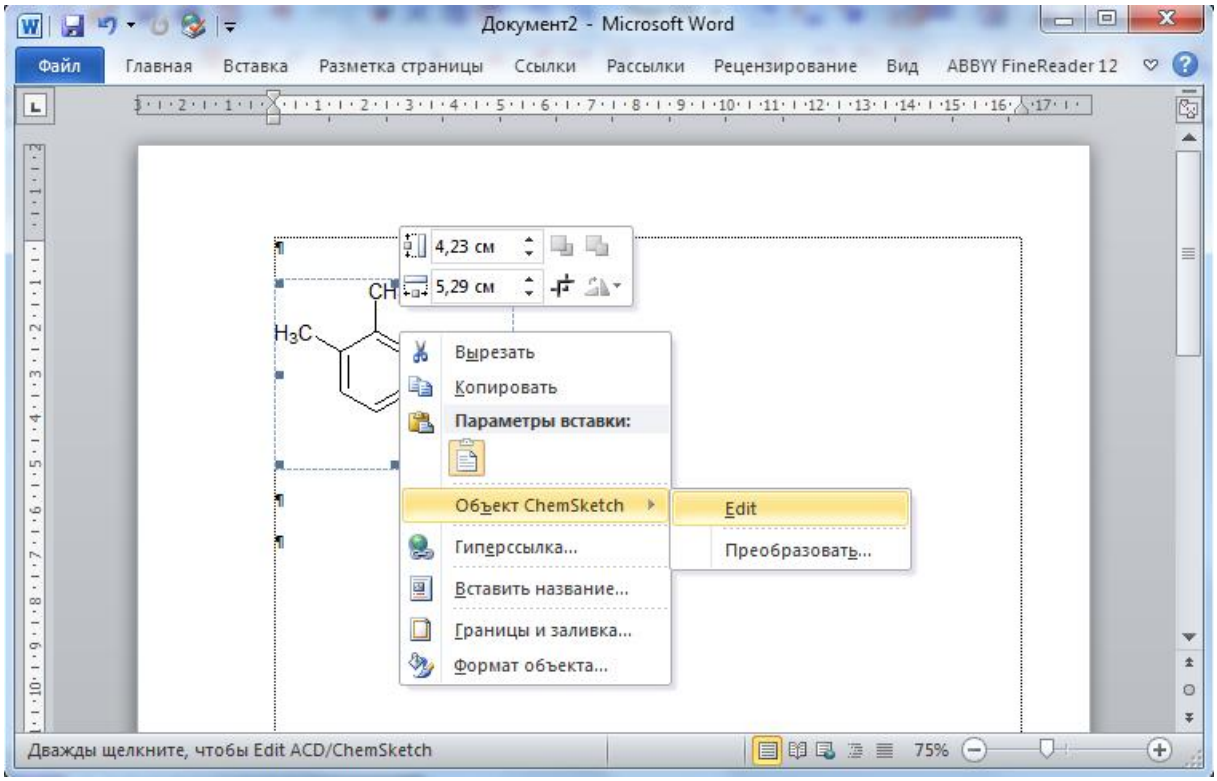


Рисунок 1.23 – Копіювання об’єктів із програми ChemDraw

Вбудовування об’єктів хімічних редакторів MarvinSketch, ChemDraw та ChemSketch до програми Word можна зробити за інший спосіб. Всі ці програми надають можливість зберегти результаті роботи у файлі одного із форматів *.jpg, *.pdf, *.tif, *.png. Надалі такі файли можуть бути вставлені до документу Word через активацію команди ***Вставка/Рисунок*** (рис. 1.24). Але слід зауважити, що такі об’єкти не підлягають подальшому редагуванню, яке іноді буває вкрай необхідним.

До програми Word можна вбудовувати графічні об’єкти із файлів формату *.cdx. Файли такого формати можуть бути створені хімічними редакторами MarvinSketch, ChemDraw та ChemSketch. Для цього спочатку треба зберегти результати роботи у хімічному редакторі у файлі формату *.cdx. Потім в програмі Word після активації команди ***Вставка/Обєкт/Створення із файлу*** у діалоговому вікні вказати шлях до файлу і натиснути кнопку ***Вставити*** (рис. 1.25).

Вбудовування даних файлів формату *.cdx до програми Word можна зробити також простим перетягуванням файлу із вікна ***Провідника Windows*** до вікна відповідного текстового документу програми Word.

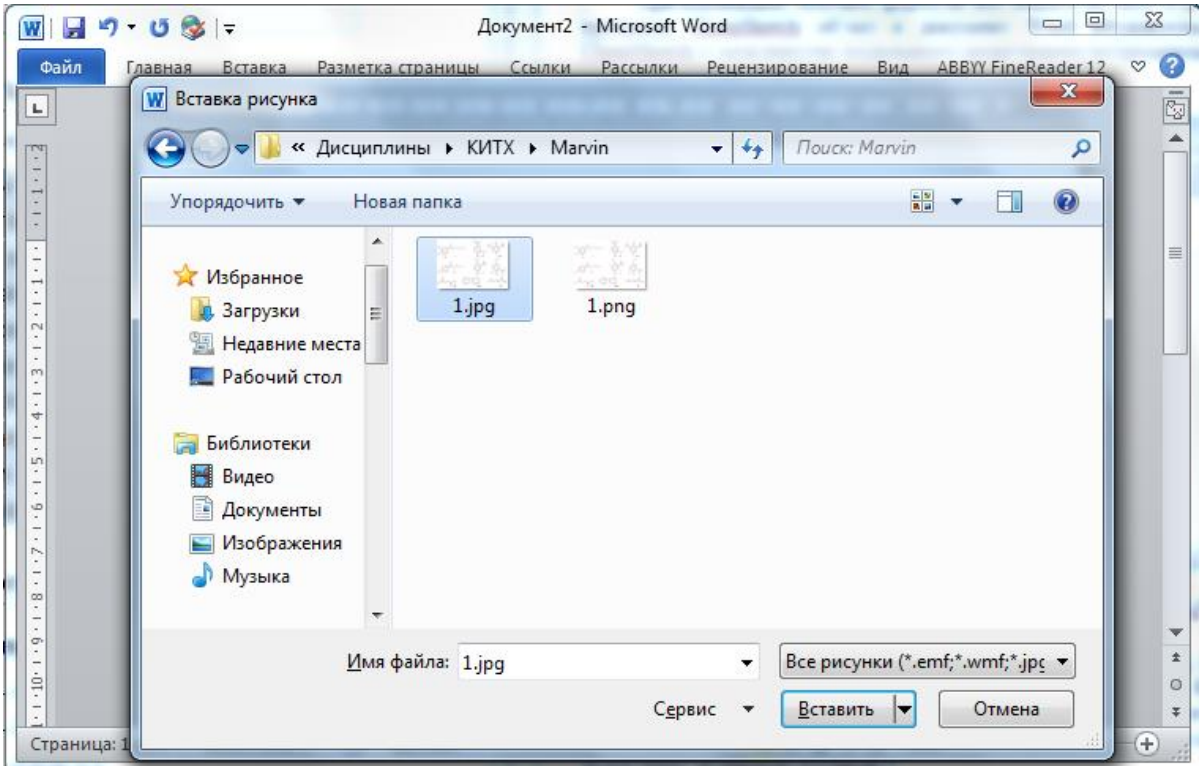


Рисунок 1.24 – Копіювання об’єктів із хімічних редакторів до програми Word у вигляді рисунків

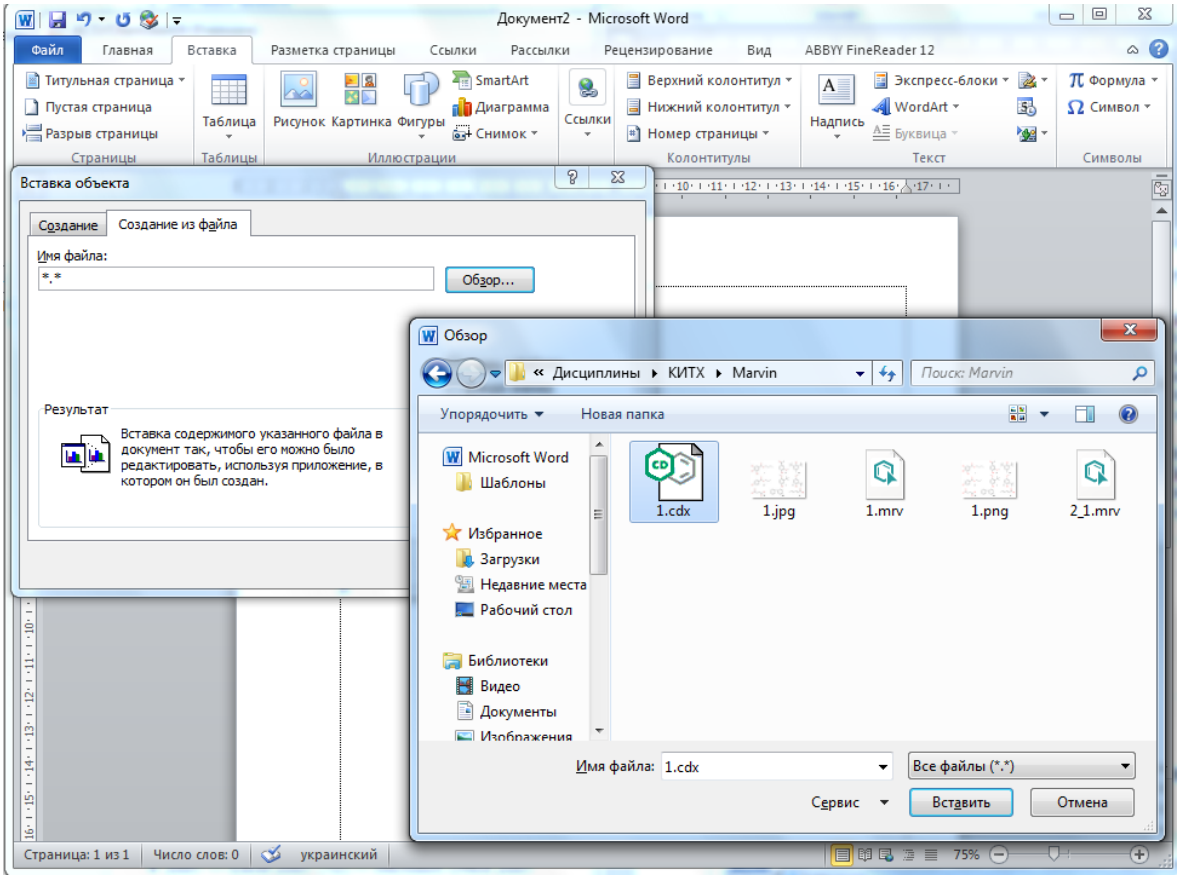


*Рисунок 1.25 – Вбудовування даних файлів формату *.cdx до програми Word*

Всі програми MarvinSketch, ChemDraw та ChemSketch надають можливість копіювати як графічні об’єкти, так і текстову інформацію. Але при копіюванні графічних об’єктів із програми MarvinSketch слід враховувати, що подальше редагування об’єкту буде неможливим. За необхідності внесення правок всі виправлення треба робити у вікні програми MarvinSketch і знову копіювати новий об’єкт до програми пакету Microsoft Office.

Microsoft Excel (повна назва Microsoft Office Excel) – табличний процесор, програма для роботи з електронними таблицями.

Завдяки тому, що лист Excel являє собою готову таблицю, Excel часто використовують для створення документів без усіляких розрахунків, що просто мають табличне представлення.

У Excel легко можна створювати різні види графіків і діаграм, що досить часто необхідно при аналізі результатів досліджень в хімії, або при статистичній обробці даних.

Excel містить багато математичних і статистичних функцій, завдяки чому його можуть використовувати школярі і студенти для розрахунків курсових, лабораторних робіт.

Excel може навіть працювати як база даних. Хоча, звичайно, до повноцінної бази даних йому далеко.

Принцип вбудовування об’єктів із програм MarvinSketch, ChemDraw та ChemSketch до листів програми Excel такий самий, як і у випадку програми Word. Але Excel дозволяє значно простіше у порівнянні з Word оформити необхідні дані у вигляді таблиць, особливо, за необхідності виконання будь-яких розрахунків і будування діаграм.

Аналогічний принцип копіювання використовується при вбудовуванні графічних об’єктів і текстової інформації із програм MarvinSketch, ChemDraw та ChemSketch до програми підготовки презентацій Microsoft PowerPoint і програми для управління базами даних (СУБД) Microsoft Access.

2 ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ В ХІМІЇ

2.1 Бази даних та хімічні каталоги в мережі Internet

Інтернет – величезне джерело практично будь-якої інформації. Існують сотні мільйонів веб-сторінок, частина яких цікава для хіміків-органіків, біоорганіків, біохіміків і спеціалістів суміжних галузей науки. Понад 80 % інформації, доступної в світовому Інтернеті, є англомовною (а в галузі науки й технології цей відсоток ще вищий). Відповідно, переважна більшість веб-ресурсів для спеціалістів у галузі хімічних наук також англомовні.

Більшість баз даних хімічного характеру (властивості молекул, методи синтезу, відповідна література й патенти тощо) в мережі Інтернет є комерційними. Найвідоміші з них – ***Beilstein, SciFinder, STN***. Безумовно, бази вільного доступу навряд чи здатні конкурувати з ними в повному обсязі й не містять такої багатої інформації. Однак в Інтернеті багато ресурсів цього класу з вільним доступом. Серед безоплатних баз слід відмітити

– **PubChem** (https://pubchem.ncbi.nlm.nih.gov) – це відкрита база даних хімії Національного інституту охорони здоров'я (NIH). Користувачи можуть розміщувати свої наукові дані в PubChem, щоб інші могли їх використовувати. База містить інформацію про хімічні структури, ідентифікатори, хімічні та фізичні властивості, біологічні дії, патенти, дані про здоров'я, безпеку, токсичність та багато інших;

– **ChemSpider** (http://www.chemspider.com) – це безкоштовна база даних хімічних структур, що забезпечує швидкий текстовий та структурний пошук серед понад 67 мільйонів структур із сотень джерел даних;

– **NIST Chemistry WebBook** (https://webbook.nist.gov/chemistry/) дозволяє вести пошук фізико-хімічних властивостей молекул (за назвою, структурою, CAS індексом, молекулярною формулою тощо); при цьому надаються термодинамічні характеристики, а в багатьох випадках навіть УФ, ІЧ і масспектри сполук, зібрані NIST в рамках Стандартної програми довідкових даних;

– **Chemical Entities of Biological Interest (ChEBI)** (https://www.ebi.ac.uk/chebi/) – це вільно доступна база молекулярних сполук, орієнтована на «малі» хімічні сполуки. ChEBI використовує номенклатуру, символіку та термінологію, затверджені IUPAC (International Union of Pure and Applied Chemistry) та NC-IUBMB (International Union of Biochemistry and Molecular Biology). Усі дані в базі є невласними або отримані з непатентованого джерела, тому база є вільнодоступною для кожного користувача (дані на цьому веб-сайті доступні за ліцензією Creative Commons (CC BY 4.0)). Кожен елемент бази даних повністю відстежується і чітко посилається на першоджерело;

– **Common Chemistry from Chemical Abstracts Service (CAS)** (http://www.commonchemistry.org/) – веб-ресурс, який містить номери реєстру

CAS для приблизно 7 900 хімічних речовин, що мають широкий загальний громадський інтерес;

– **SDBS – Spectral Database System For Organic Compounds** (https://sdbs.db.aist.go.jp/sdbs/cgi-bin/cre_index.cgi) є інтегрованою системою спектральних даних для органічних сполук;

– **eMolecules** (https://www.emolecules.com/) дозволяє знайти фізико-хімічні дані молекул за структурою, а також інформацію про виробників цих сполук (у базі понад 7 млн структур);

– **ChemID*plus* A TOXNET DATABASE of National Library of Medicine** (https://toxnet.nlm.nih.gov/) – ресурс для пошуку баз даних про токсикологію, небезпечні хімічні речовини, екологічне здоров'я та токсичні викиди (16 грудня 2019 року інформація Національної бібліотеки медицини (NLM) TOXNET перенесена на PubChem, PubMed та Bookshelf);

– **Organic Syntheses** (http://www.orgsyn.org/) дозволяє отримати методики синтезу й використання в синтезі сполук як за ключовими словами, так і за структурою (із застосуванням плагіну ChemDraw);

– **Merck** (https://www.sigmaaldrich.com/european-export.html) – онлайновий хімічний каталог, який дозволяє здійснювати швидкий структурний і текстовий пошук хімічних сполук і який містить багато корисної для хіміків інформації.

2.1.1 Ресурс PubChem

Ресурс PubChem (https://pubchem.ncbi.nlm.nih.gov) являє собою базу даних ресурсів Національного центру інформації з біотехнології (The National Center for Biotechnology Information). Дозволяє здійснювати швидкий структурний пошук а також прогнозування біологічної активності органічних сполук. Кілька оновлень веб-інтерфейсу PubChem було зроблено у березні 2019 року. Оновлення включають новий вигляд для більшості сторінок вмісту PubChem, зокрема, результати пошуку, записи про історію та біоаналіз, вигляд патентів тощо. Крім того, домашня сторінка PubChem та пошук були повністю оновлені. Було створено єдиний інтерфейс хімічного пошуку, який підтримує пошук тексту у всіх колекціях даних PubChem, а також пошук структури та молекулярних формул. Цей єдиний інтерфейс пошуку став новою домашньою сторінкою PubChem.

Інформаційний пошук в цієї базі можна здійснювати за 1D-ідентифікаторами (систематична назва, назва за номенклатурою ІЮПАК, лінійні коди, реєстраційні номери сполуки), за точною структурною формулою, яка вводиться за допомогою відповідного апплету (рис. 2.1). Також для вказаної структури виконується пошук літературних джерел (статті, книги, патенти), фізико-хімічних даних, структурної, спектральної інформації, тощо.

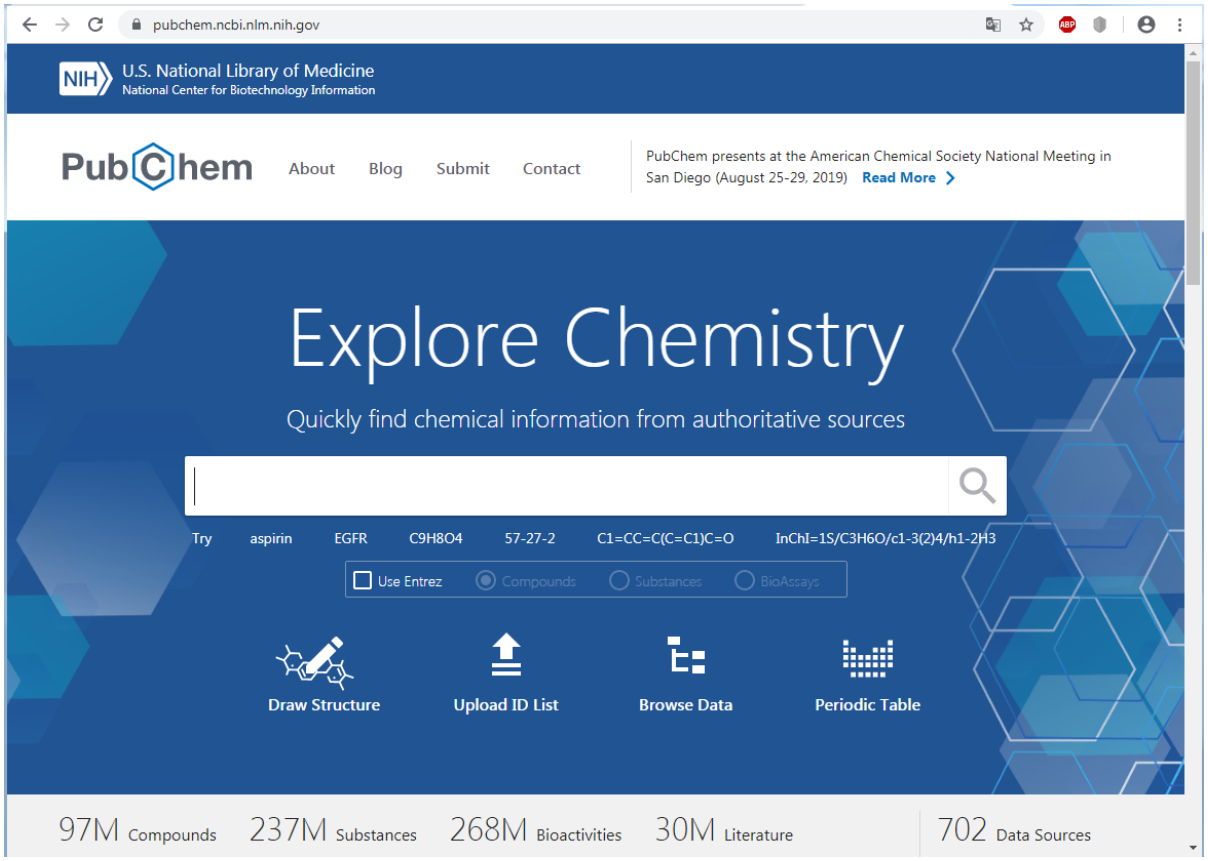


Рисунок 2.1 – Вікно пошуку ресурсу PubChem

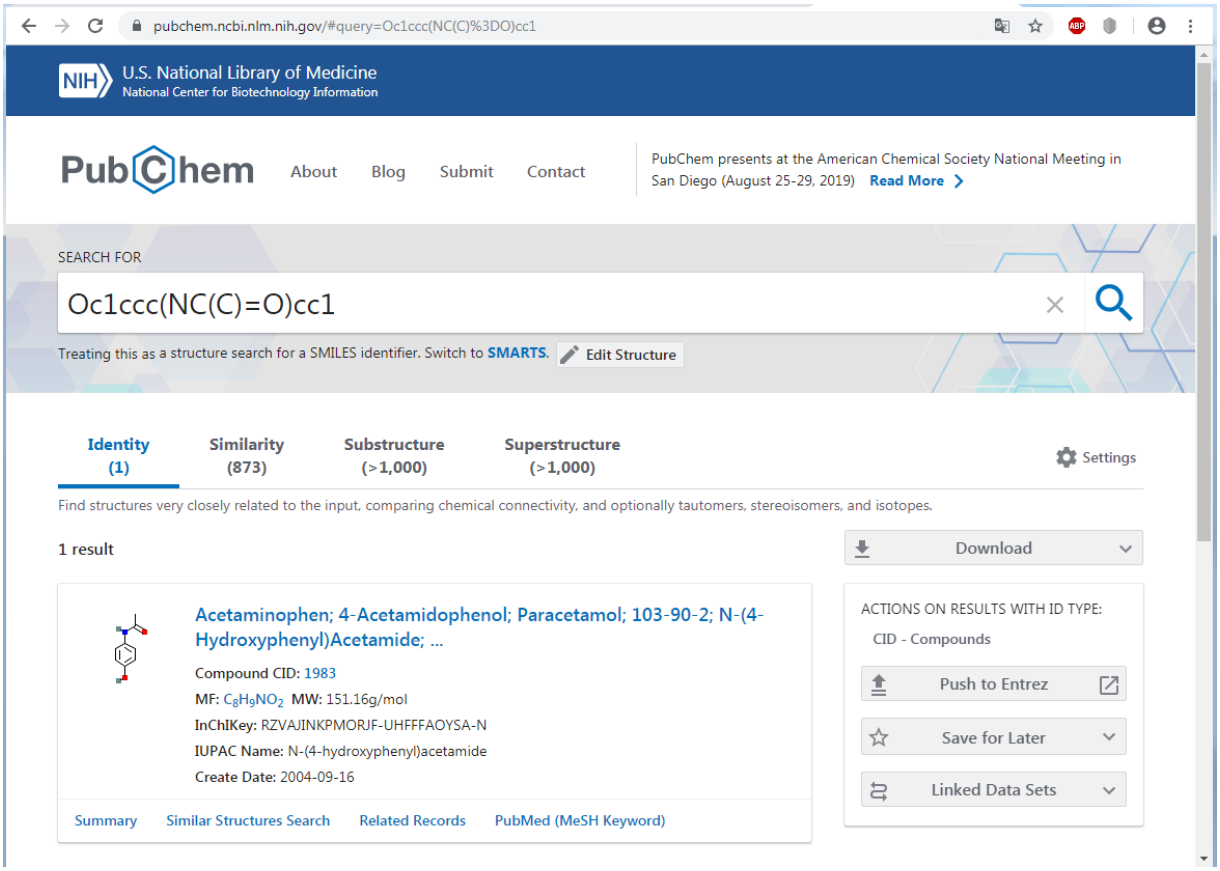


Рисунок 2.2 – Результати пошуку за SMILES кодуванням

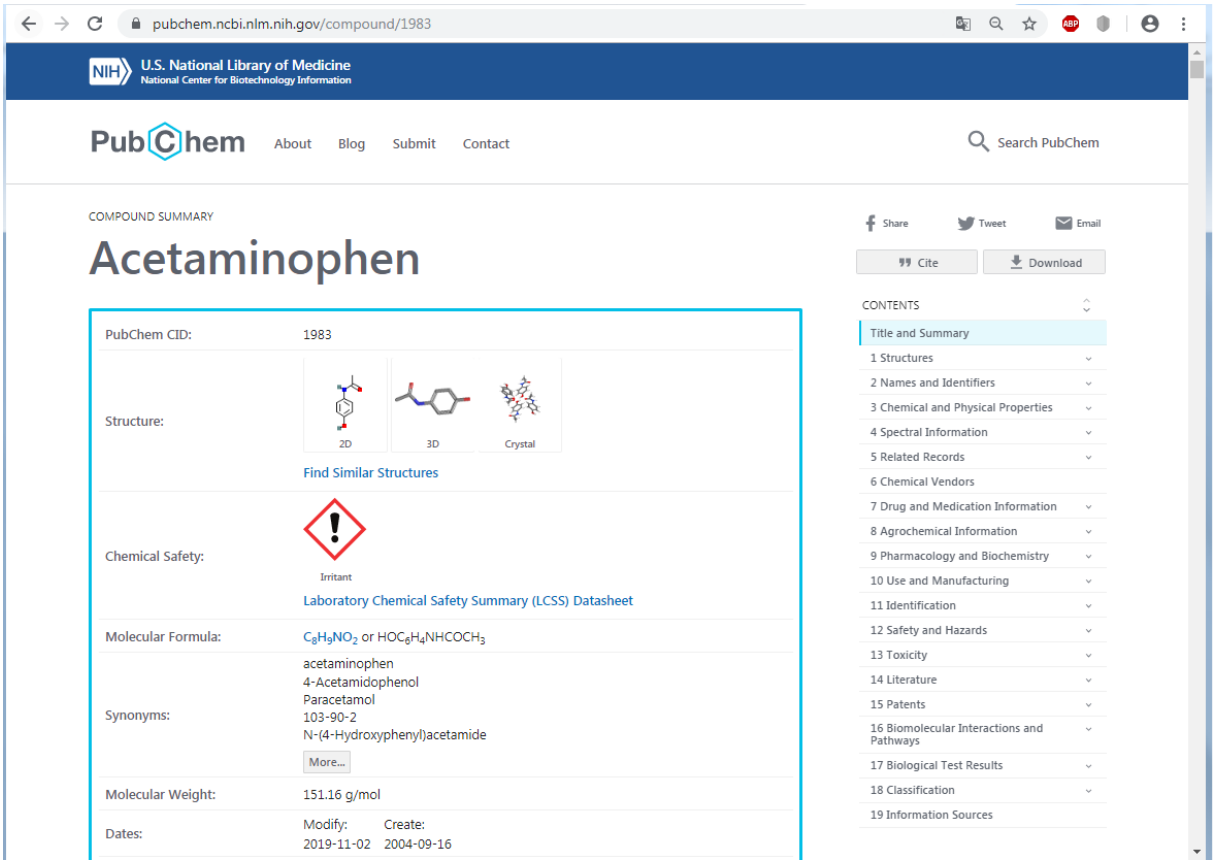


Рисунок 2.3 – Інформація про хімічну сполуку, що міститься за посиланням ***Summary***

Результати пошуку даних про хімічну речовину (див. рис. 2.1 – активна опція ***Compounds*** нижче пошукової строки) мають вигляд, наведений на рисунку 2.2. Наприкінці цієї сторінці за посиланням ***Summary*** знаходиться докладна інформація про дану сполуку із вказівкою джерела даних за наступними категоріями (рис. 2.3):

1. структури;
2. імена та ідентифікатори;
3. хімічні та фізичні властивості;
4. спектральна інформація;
5. суміжні записи;
6. постачальники хімічних сполук;
7. інформація про ліки та медикаменти;
8. агрохімічна інформація;
9. фармакологія та біохімія;
10. використання та виготовлення;
11. ідентифікація;
12. безпека та небезпека;
13. токсичність;
14. література;
15. патенти;
16. біомолекулярні взаємодії та шляхи;

17. результати біологічних випробувань;
18. класифікація;
19. джерела інформації.

2.1.2 Ресурс ChemSpider

Ресурс ChemSpider (http://www.chemspider.com) являє собою безкоштовну базу даних, яка відноситься до ресурсів Королівського хімічного товариства. Дозволяє здійснювати швидкий структурний і текстовий пошук, містить більше ніж 77 мільйонів записів. Має зв'язок із 276 базами даних. Нова інформація до бази ChemSpider додається майже щоденно. На даному етапі вона інтегрована з процесом публікації в системі RSC Publishing, тому інформація про нову сполуку додається відразу після публікації статті.

Інформаційний пошук можна здійснювати за 1D-ідентифікаторами (систематична назва, назва за номенклатурою ІЮПАК, лінійні коди, реєстраційні номери сполуки), за точною структурною формулою, яка вводиться за допомогою відповідного апплету (рис. 2.4).

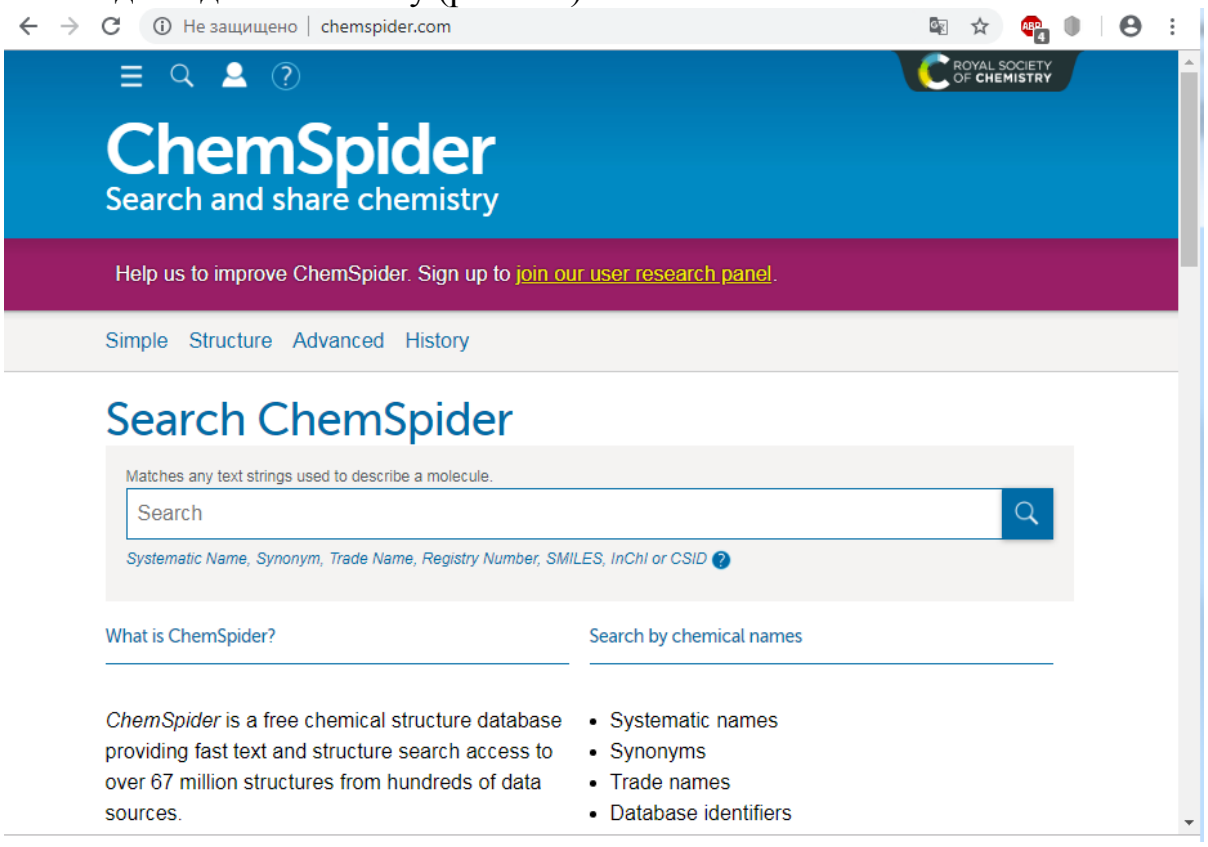


Рисунок 2.4 – Вікно пошуку ресурсу ChemSpider

Ресурс ChemSpider містить наступну інформацію про дану хімічну сполуку: назва за систематичною номенклатурою, відомі ідентифікатори та реєстраційні номери інших важливих електронних ресурсів (наприклад, CAS

номер), фізичні та хімічні властивості сполуки, молекулярну структуру, спектральні дані, інформацію про токсичність, фармакологічні властивості, перелік бібліографічних посилань на статті у сучасних наукових літературних джерелах, патентах.

Результати пошуку даних парацетамолу, виконаного за SMILES кодуванням, наведені на рисунку 2.5.

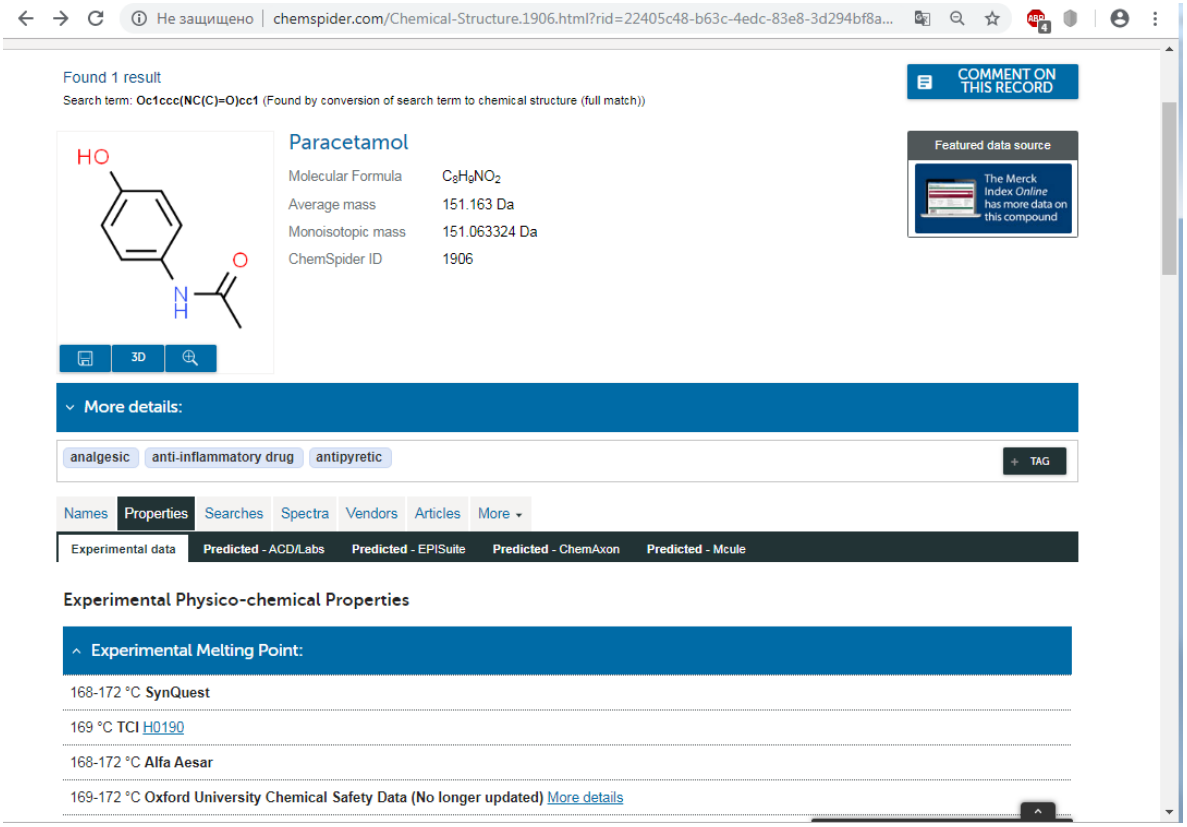


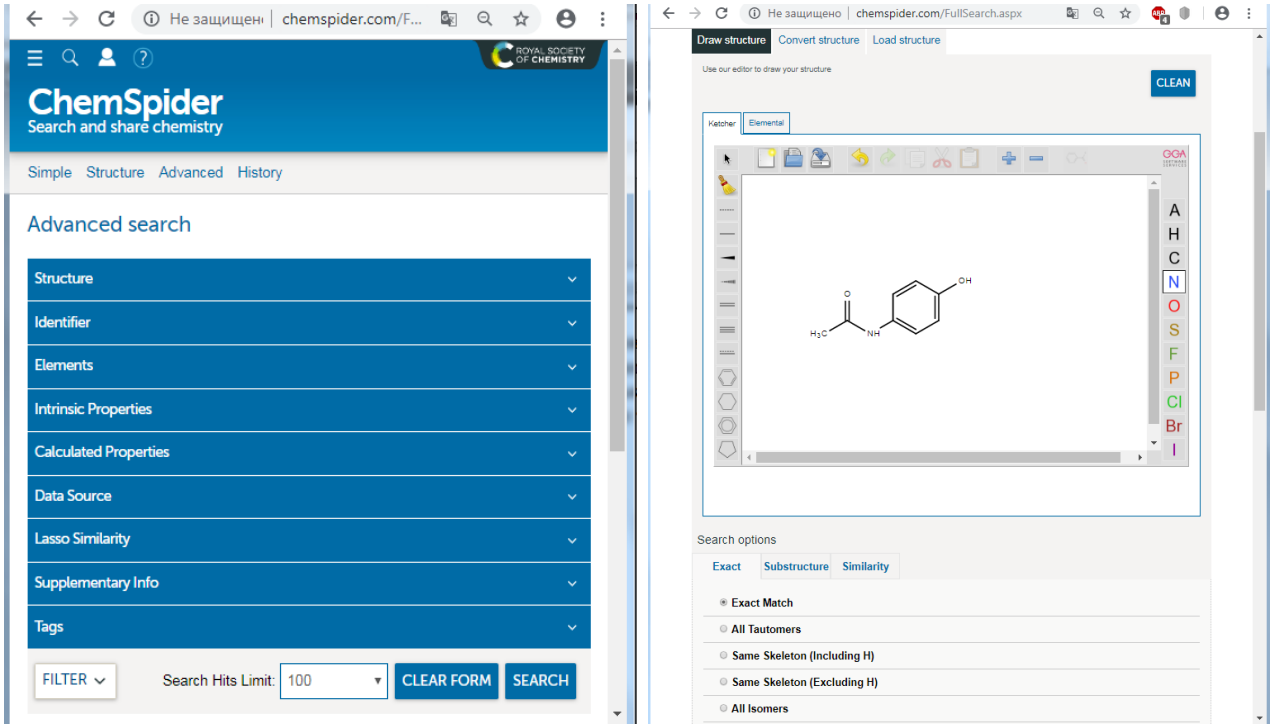
Рисунок 2.5 – Результати пошуку даних парацетамолу

Дані про сполуку наводяться на окремих вкладках: ***Імена (Names), Властивості (Properties), Пошук (Searches), Спектри (Spectra), Виробники (Vendors), Статті (Articles)*** та ***Інші (More)*** (див. рис. 2.5).

Розширений пошук (рис. 2.6, *а*) можна проводити за відомою структурною формулою за допомогою вбудованого апплету (рис. 2.6, *б*), ідентифікаторами, елементним складом, властивостями (експериментальними або розрахованими кількісними фізико-хімічними характеристиками), типом даних, типом джерел тощо.

На вкладці ***Історія*** (***History***) наводиться перелік пошуків, виконаних у даній системі.

Пошук даних в системі ChemSpider може бути також виконаний на мобільних пристроях за допомогою безкоштовних додатків для iOS (iPhone / iPod / iPad) і для Android.



а) *б)*

а – вікно розширеного пошуку;
б – вікно пошуку за відомою структурною формулою
Рисунок 2.6 – Розширений пошук у системі ChemSpider

2.1.3 Ресурс e-Molecules

Ресурс e-Molecules (https://www.emolecules.com/) являє собою онлайновий ресурс пошуку інформації у власній базі даних ресурсу та провідних хімічних каталогах і базах даних властивостей хімічних сполук. Дозволяє здійснювати швидкий структурний і текстовий пошук.

eMolecules – це комерційний проект, який дозволяє купувати сполуки по всьому світу за допомогою декількох клацань, а потім легко відстежувати їх, щоб переконатися, що сполуки надходять у потрібному виді, чистоті та термінах. У базі даних eMolecules відображаються лише дані про сполуки, які є в наявності або можуть бути синтезовані (віртуальні сполуки обробляються окремо).

Ця пошукова система дозволяє виконувати пошук через підструктуру, подібність або точний пошук, використовувати файл SD, SMILES, хімічну назву, номер CAS, номер каталогу або ідентифікаційний номер eMolecules.

Ресурс містить зручний аплет для здійснення структурного пошуку за структурною формулою хімічної сполуки або її фрагмента (рис. 2.7).

Результати пошуку наведено на рисунку 2.8. Для перегляду докладної інформації про хімічну сполуку (див. рис. 2.8, посилання ***View Compound Info***) треба зареєструватися в системі.

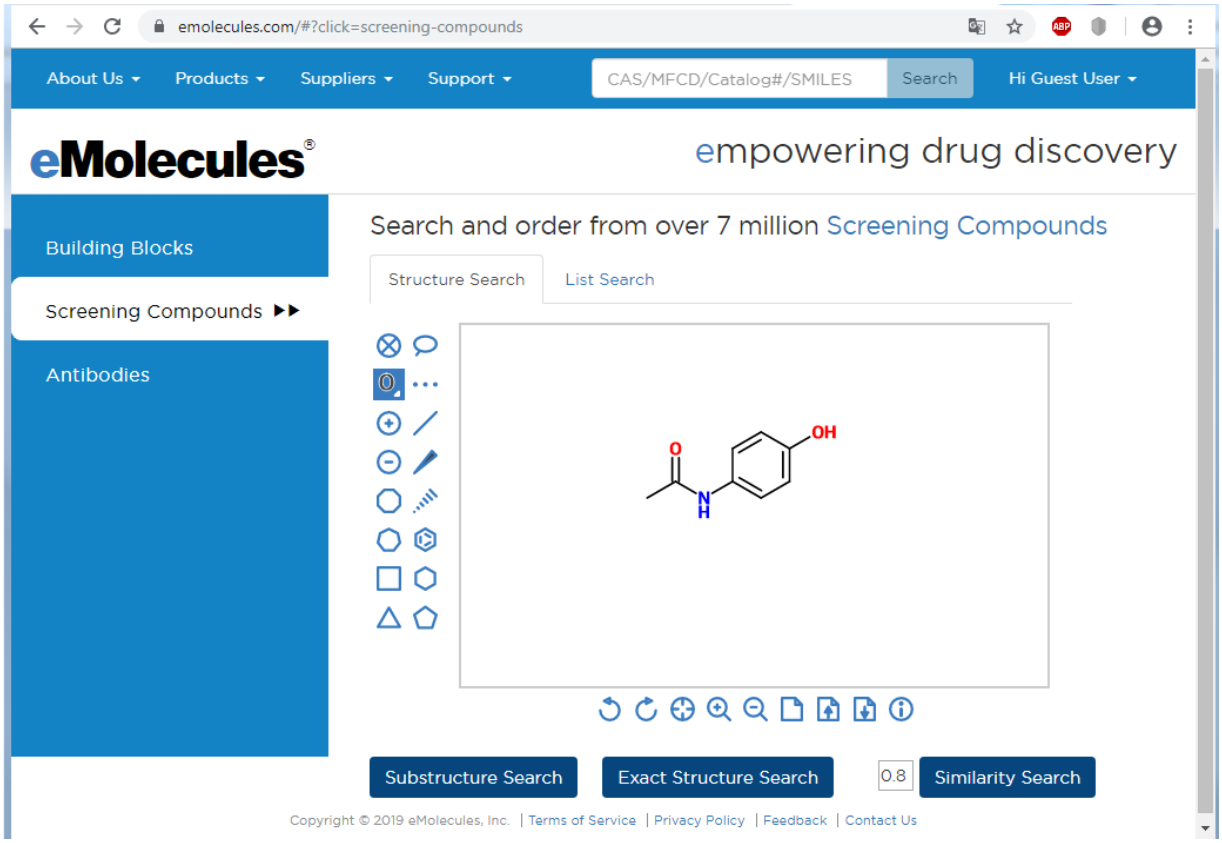


Рисунок 2.7 – Вікно пошуку хімічної сполуки за структурною формулою на ресурсі eMolecules

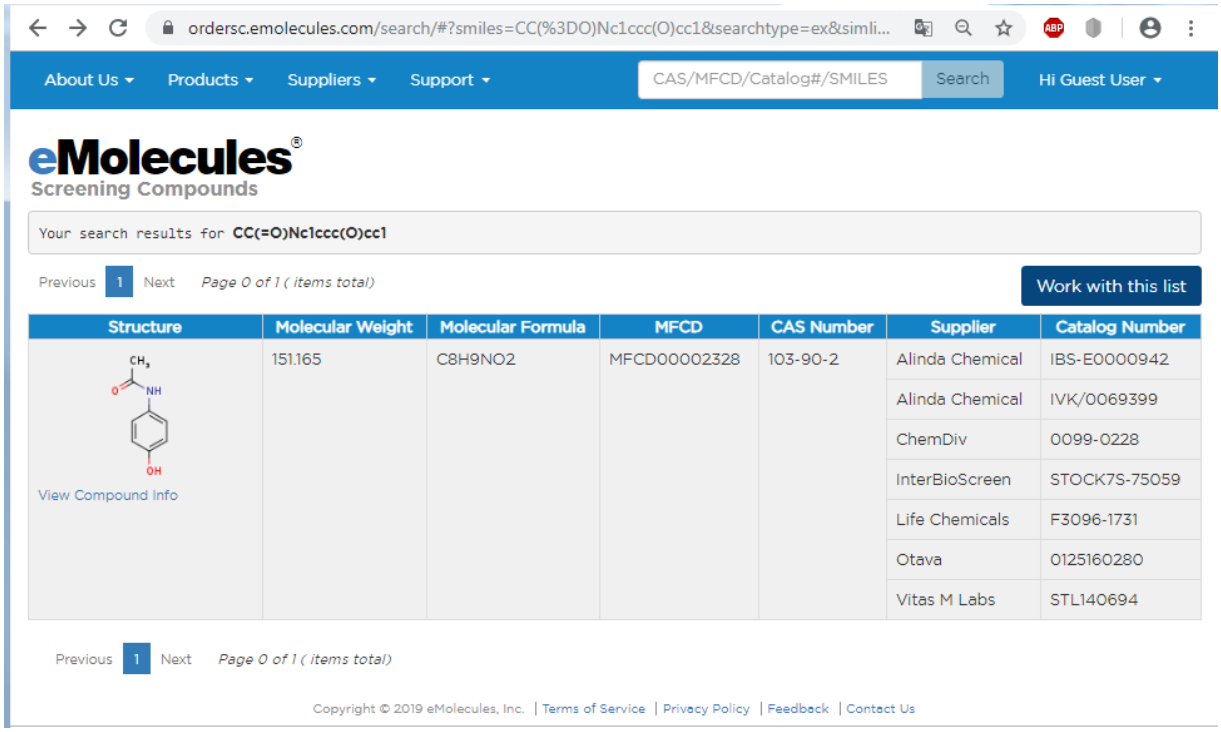


Рисунок 2.8 – Результати пошуку в системі eMolecules

2.1.4 Ресурс COMMON CHEMISTRY (CAS)

База даних COMMON CHEMISTRY (http://www.commonchemistry.org/) містить номер реєстру CAS®, хімічні назви (як формальні, так і загальні), молекулярні формули та структури для близько 7900 хімічних речовин, що мають широкий загальний громадський інтерес. До бази включено речовини, які мають глобальне комерційне використання або важливість, і їх цитували 1000 і більше разів у базах даних CAS. Прикладами речовин, які входять, є аспірин, біотин, пероксид бензоїлу, борна кислота та ін.

База даних COMMON CHEMISTRY також включає всі 118 елементів Періодичної таблиці, хоча не всі елементи можуть відповідати межі 1000 посилань.

Посилання на записи Вікіпедії в базі, якщо вони є, надаються базі проектом WikiProject Wikipedia у співпраці зі Службою хімічних рефератів.

Пошук речовини в базі COMMON CHEMISTRY можна вести за номером реєстру CAS або за назвою хімічної сполуки. Пошук назви можна здійснювати за точною назвою або за її фрагментом (рис. 2.9).

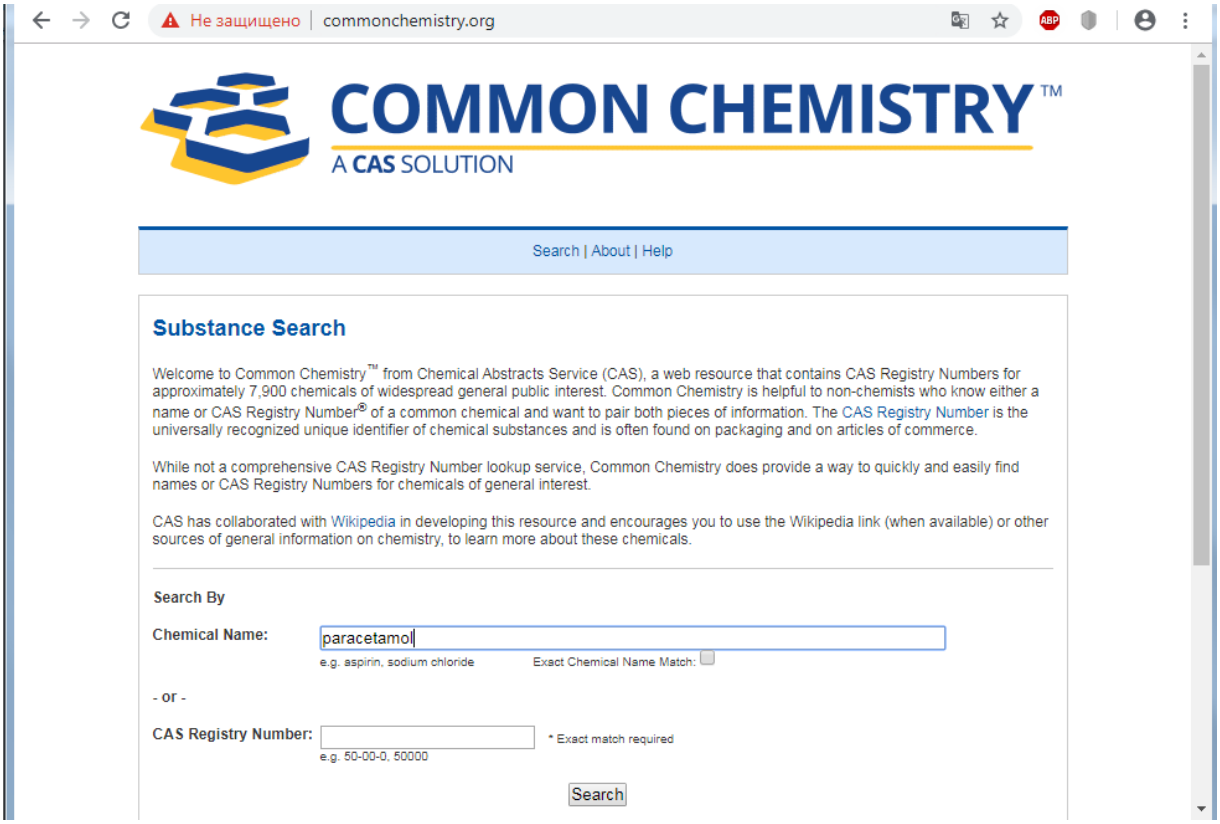


Рисунок 2.9 – Вікно пошуку в системі COMMON CHEMISTRY

На запит надається перелік всіх можливих назв сполуки, її структурна формула та номер реєстру CAS (рис. 2.10).

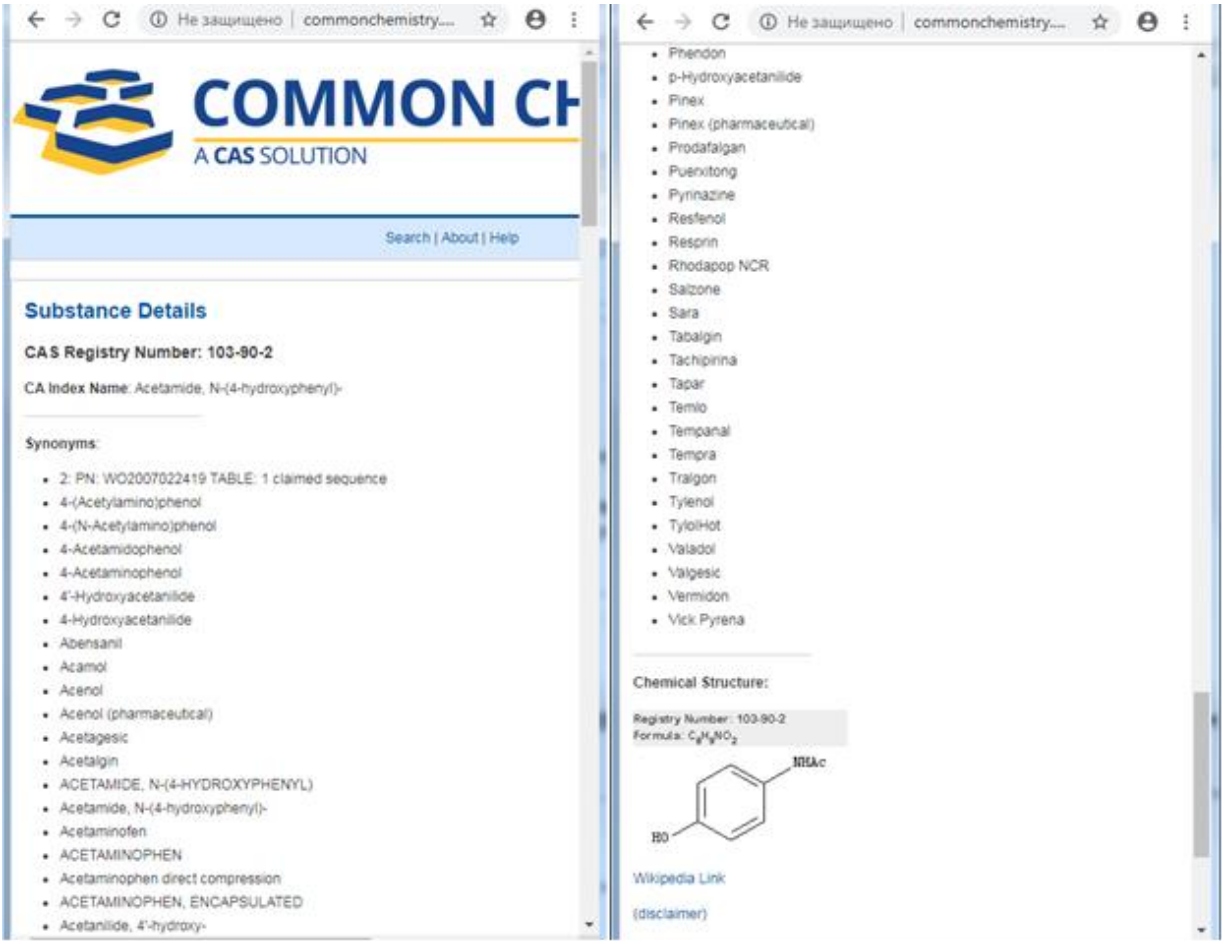


Рисунок 2.10 – Результати пошуку в системі COMMON CHEMISTRY

2.1.5 Ресурс Organic Syntheses

Ресурс Organic Syntheses – це електронна версія наукового журналу «Organic Syntheses», який видається з 1921 року. Organic Syntheses (http://www.orgsyn.org/) надає хімічній спільноті детальні, надійні та ретельно перевірені процедури синтезу органічних сполук. Деякі процедури описують практичні методи приготування конкретних цікавих сполук, а інші процедури ілюструють важливі синтетичні методи із загальною корисністю. Кожна процедура, наведена в базі, написана значно детальніше порівняно з типовими експериментальними процедурами в інших журналах.

Перевагою цієї бази є те, що кожна реакція з її характеристичними даними була повторена кілька разів і ретельно "перевірена" на відтворюваність у лабораторіях членів Ради редакторів цього ресурсу.

Ресурс Organic Syntheses публікується некомерційною корпорацією Organic Syntheses, Inc.

До 1998 року «Органічні синтези» виходили лише як щорічний друкований випуск. Але з 1998 року всі томи журналу стали доступними на веб-

сайті з відкритим доступом, а нові статті публікуються в Інтернеті, як тільки вони приймаються.

На головній сторінці ресурсу Organic Syntheses доступний швидкий пошук за структурою, за назвою або за вихідними даними методики синтезу (рис. 2.11).

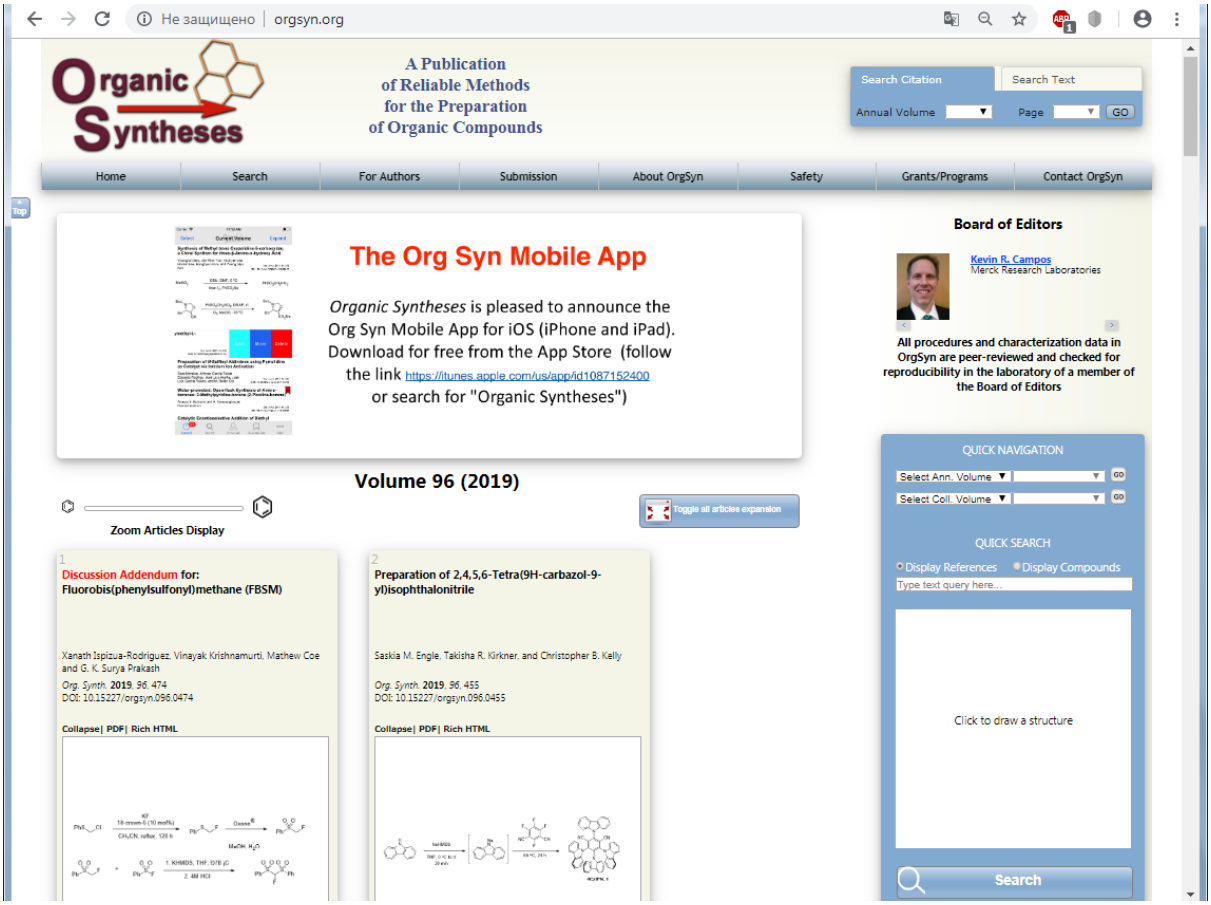


Рисунок 2.11 – Головна сторінка Organic Syntheses із вікном пошуку

При активації команди ***Display References*** на запит користувача (наприклад, за структурною формулою, рис. 2.12) система надає перелік синтезів, в яких дана або аналогічна за структурою речовина являється або вихідною сполукою, або кінцевим продуктом, або інтермедіатом (рис. 2.13).

При активації команди ***Display Compounds*** на запит користувача система надає перелік структур хімічних сполук, які мають схожу структурну формулу або містять в своїй структурі фрагмент запиту (рис. 2.14). Для кожної сполуки при цьому вказано її номер реєстру CAS (якщо він є) і надається посилання на відповідні синтези, в яких дана сполука являється або вихідною сполукою, або кінцевим продуктом, або інтермедіатом.

Кожну методику синтезу, надану системою, можна переглянути у форматі PDF або HTML. Кожна методика має ідентифікаційний номер DOI (рис. 2.15).

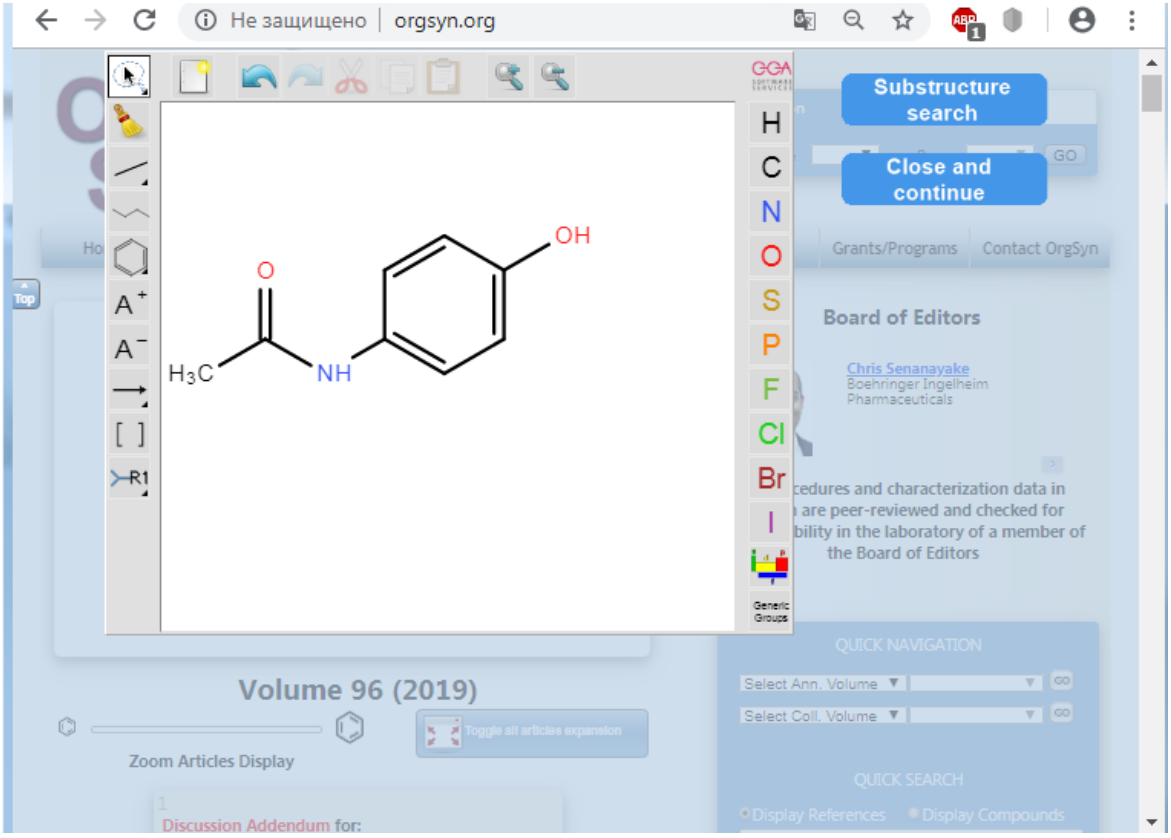


Рисунок 2.12 – Запит пошуку за структурною формулою методик синтезу аналогів парацетамолу в системі Organic Syntheses

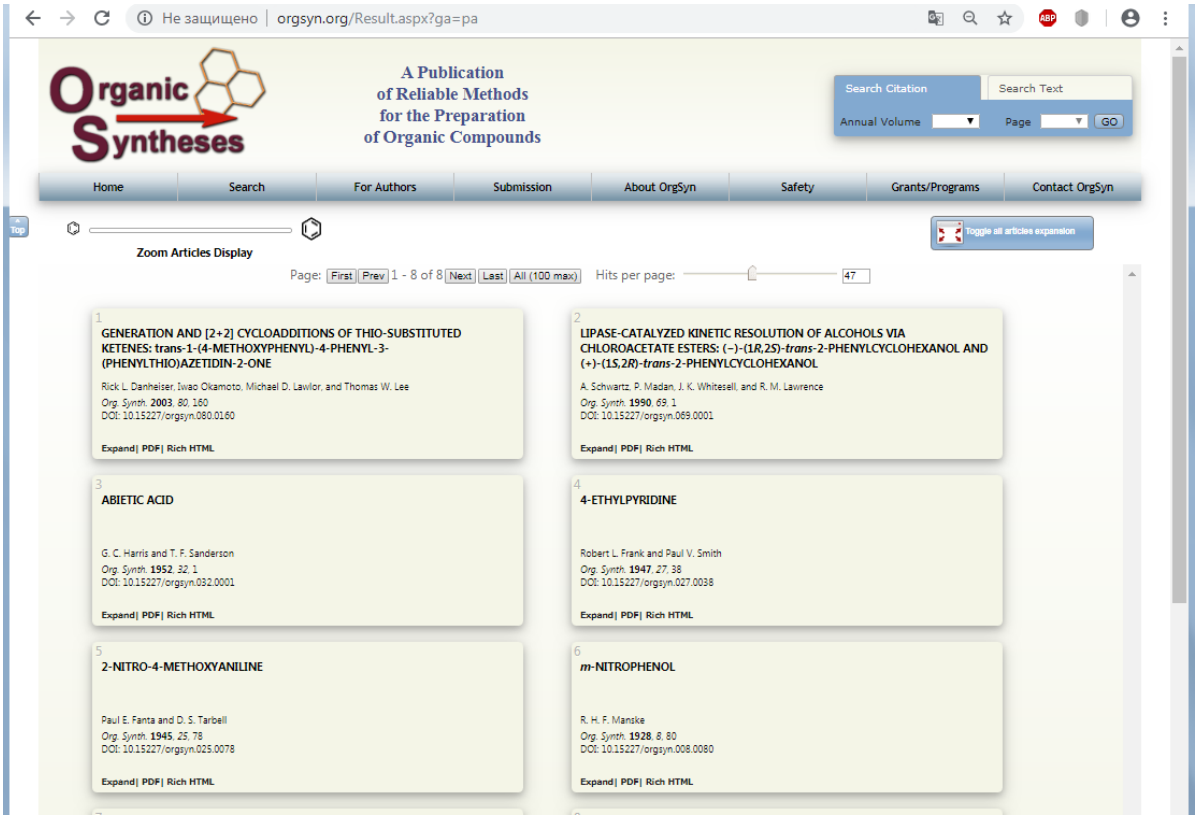


Рисунок 2.13 – Результати пошуку в Organic Syntheses при активації команди Display References

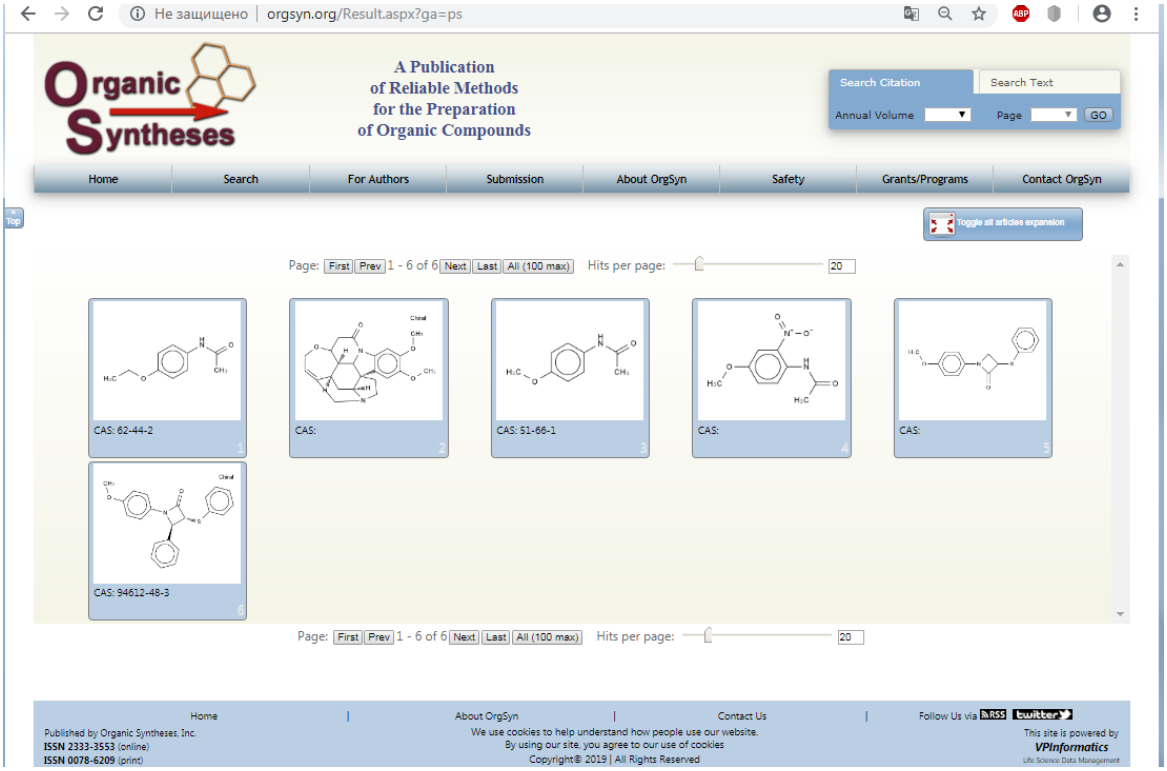


Рисунок 2.14 – Результати пошуку в Organic Syntheses при активації команди Display Compounds

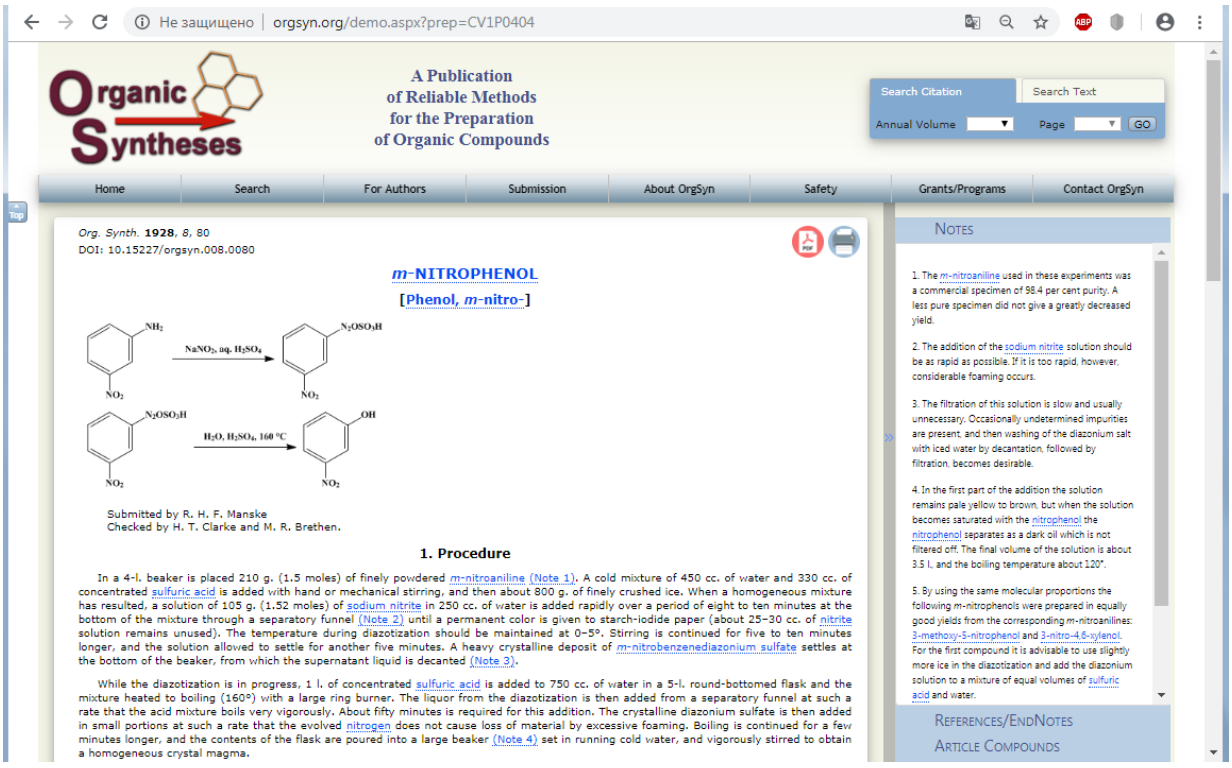


Рисунок 2.15 – Приклад методики синтезу в Organic Syntheses

2.2 Електронні ресурси хімічної наукової періодики

Практично всі більш-менш серйозні наукові журнали сьогодні доступні через Інтернет, зокрема, й повнотекстові. Існують також онлайнові журнали, які не мають друкованої версії. Для пошуку публікацій у фахових виданнях можна використовувати різні шляхи, але є певні алгоритми, які дозволяють з мінімальними витратами часу знайти необхідну інформацію.

2.2.1 Наукова періодика України

В Україні створено Реєстр наукових фахових видань України (http://nfv.ukrintei.ua/), який дозволяє знайти всі необхідні дані про фахові наукові видання. На головній сторінці Реєстру доступний ***швидкий пошук*** з можливістю уточнення за наступними категоріями (рис. 2.14):

– категорія видання (категорія А, Б або В);
– вид видання (журнал або збірник);
– засновник;
– галузь науки (треба вибрати із переліку);
– спеціальність (треба вибрати із переліку);
– мова видання (треба вибрати із переліку);
– тип видання (друковане, онлайн, на CD, виключно онлайн, виключно друковане).

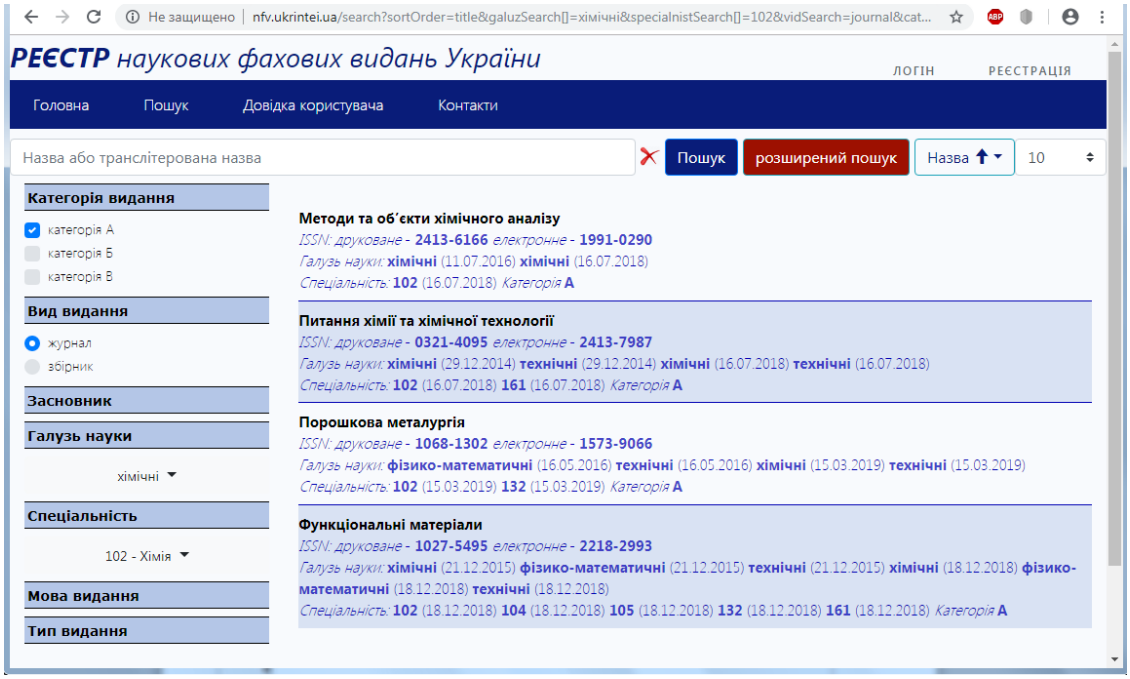


Рисунок 2.16 – Вікно пошуку Реєстру наукових фахових видань України

Розширений пошук дозволяє уточнити певні критерії пошуку і значно скоротити перелік записів, що надаються при запиті (рис. 2.17).

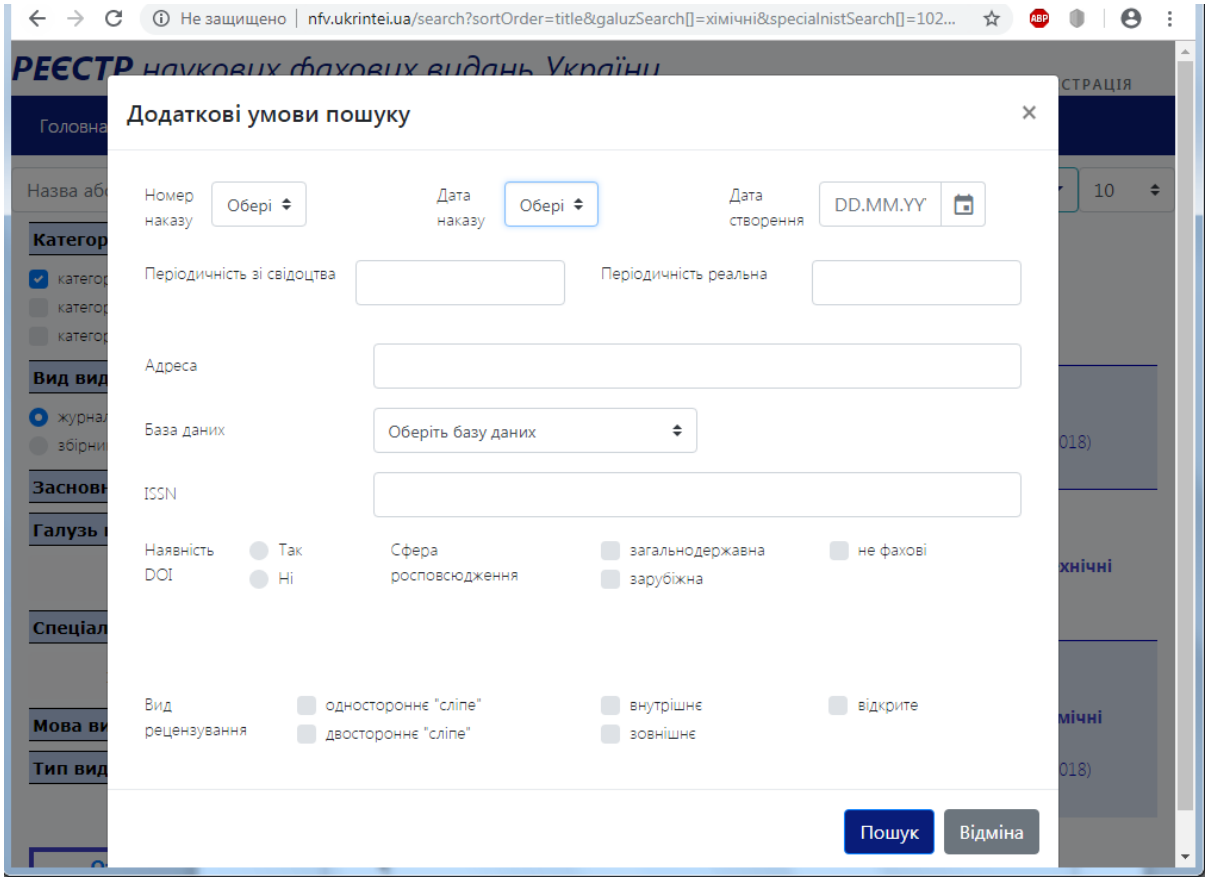


Рисунок 2.17 – Вікно розширеного пошуку Реєстру наукових фахових видань України

На сторінці опису журналу або збірника наводиться наступна інформація (рис. 2.18):

- категорія видання;
- ISSN видання;
- вихідні дані;
- галузь науки;
- спеціальність;
- періодичність видання;
- мова статей і анотацій;
- посилання на домашню сторінку;
- посилання на архів номерів;
- тематика;
- міжнародні бази та каталоги, які індексують видання;
- контактні дані.

Правила оформлення публікацій для подання до редакційної колегії, реферати або повні тексти статей можна знайти за посиланням на домашній сторінці відповідного видання.

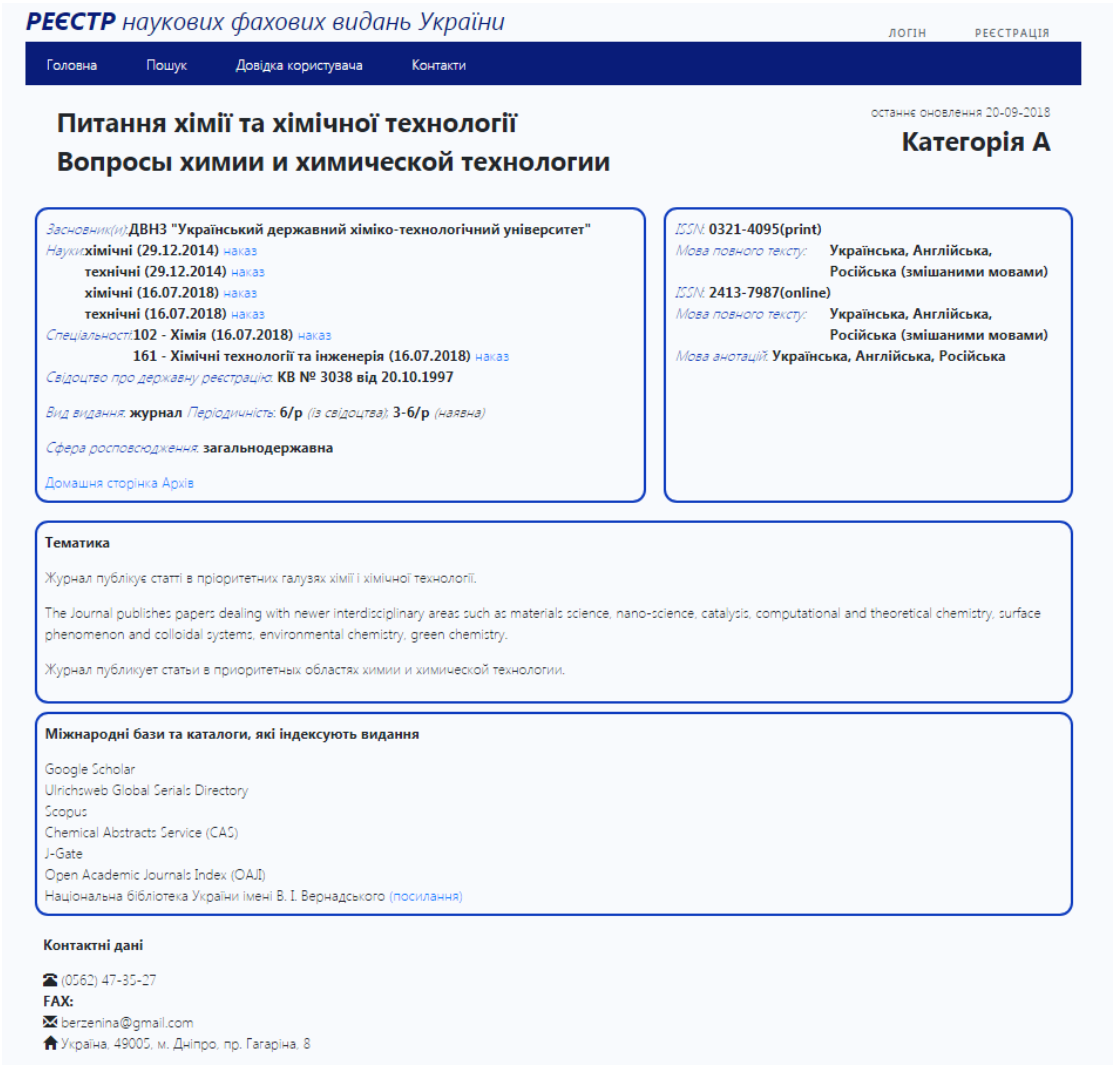


Рисунок 2.18 – Інформація про журнал «Питання хімії та хімічної технології» з Реєстру наукових фахових видань України

Перелік журналів «***Наукової періодики України»*** також представлений за посиланням http://journals.uran.ua/search/category/261. Це загальнодержавна технологічна платформа, яка забезпечує процеси редакційного опрацювання, публікації та післяпублікаційної підтримки наукових періодичних видань України. Ресурс розвивається на засадах партнерства видавців та бібліотек України. Платформа надає посилання на офіційну сторінку відповідного видання. Можливий пошук видань за певною галуззю знань. Наприклад, ця платформа надає перелік з 10 видань хімічного напряму.

Перелік українських видань, зокрема, в галузі хімічних наук, які входять до вітчизняних та міжнародних наукометричних, реферативних баз даних, наведений на сайті ***наукової бібліотеки Чернігівського національного технічного університету*** (http://library2.stu.cn.ua/) в розділі ***На допомогу науковцю*** за посиланням ***Українські наукові фахові видання***. Для кожного видання платформа надає посилання на офіційну сторінку видання або на сторінку видання на сайті Національної бібліотеки ім. В.І. Вернадського.

Національна бібліотека ім. В.І. Вернадського надає можливість знайти будь-яке видання наукової періодики України за посиланням (http://nbuv.gov.ua/). В розділі ***Науковий пошук*** (http://nbuv.gov.ua/node/1539/) за посиланням ***Наукова періодика України*** знаходиться електронний архів наукових періодичних видань України, який містить 2726 журналів і 1042786 повних текстів статей. Можливий пошук видань за різними критеріями.

Основні профільні українські видання, що мають повнотекстові онлайнові версії, – це **Ukrainica Bioorganica Acta** (http://www.bioorganica.org.ua/), **Український біохімічний журнал** (http://ua.ukrbiochemjournal.org/), **Видавничий дім "Академперіодика** (https://akademperiodyka.org.ua/uk).

В Україні створено **Open Ukrainian Citation Index (OUCI)** (http://ouci.dntb.gov.ua/). Це пошукова система і база даних наукових цитувань, які надходять від усіх видань, що використовують сервіс Cited-by від Crossref та підтримують Initiative for Open Citations. OUCI покликаний спростити пошук наукових публікацій, привернути увагу редакцій до проблеми повноти та якості метаданих українських наукових видань, покращити представлення українських наукових видань у спеціалізованих пошукових системах.

2.2.2 Наукова періодика світу

На даний час, практично всі наукові журнали в тій чи іншій формі представлені в Інтернет. Основні варіанти представлення журналу:

– загальна інформація про журнал, бібліографія (зміст за томамі і номерами);

– загальна інформація про журнал, бібліографія, реферати;

– загальна інформація про журнал, бібліографія, реферати, повні тексти статей (найчастіше в форматі PDF, який вимагає використання вільно поширюваного програмного забезпечення Adobe Acrobat Reader).

Загальна інформація про журнал, а також бібліографія і реферати найчастіше надаються безоплатно, а повні тексти статей (за винятком журналів відкритого доступу) – обмежено (в межах передплати або за окрему плату).

Інформація про журнали найчастіше розміщується на сайтах видавництв, в рамках яких ці журнали виходять у світ, або на сайтах наукових товариств і організацій відповідного спрямування. Тому найпростіший спосіб знайти в Інтернеті науковий журнал (якщо він там представлений) – звернутися до сервера організації, яка його видає.

Однак, якщо видавництво (або його Інтернет-адреса) невідомо, інформацію про присутність журналу в Інтернеті і відповідну веб-адресу можна отримати через довідкові сторінки періодичних видань в Інтернеті. Цими ж довідниковими сторінками (покажчиками або пошуковими системами) можна скористатися, якщо треба знайти видання з певної тематики.

До політематичних баз даних та пошукових систем, що надають безоплатний доступ до своєї інформації, відносяться:

– **arXiv (Cornell University Library)** (https://arxiv.org/) надає безоплатний доступ до науково-технічних видань з фізики, математики, біології, комп'ютерних технологій;

– **BASE (Bielefeld Academic Search Engine)** (https://www.base-search.net/) – система, що спеціалізується на пошуку наукових документів відкритого доступу в Інтернеті (оператором BASE є бібліотека університету Білефельд, Німеччина);

– **Bentham Open** (https://benthamopen.com/index.php) – безоплатний електронний ресурс з різної тематики, у вільному доступі понад 100 журналів;

– **Electronic Journals Library** (Elektronische Zeitschriftenbibliothek EZB) (http://ezb.uni-regensburg.de/?lang=en) – електронна бібліотека журналів при бібліотеці університету Регенсбурга, надає посилання на майже 60 тис. безкоштовних журналів з різних галузей науки, техніки;

– **Google Book Search** (https://books.google.com/) охоплює колекцію оцифрованих книг, база даних корисна для перевірки та пошуку цитувань;

– **Google Scholar** (https://scholar.google.com/) містить анотації книг, реферати, статті з різної тематики (у списку результатів пошуку відображаються тільки цитати з відповідних документів, а доступ до повних текстів статей не забезпечується);

– **HighWire Press** (Stanford University (США) (http://highwire.stanford.edu/) містить реферати та повні тексти статей з журналів з різних галузей знань (можливий перегляд журналів в алфавітному порядку, а також за видавництвами; є інформація про наявність вільного доступу до окремих випусків або до журналу в цілому);

– **Hindawi Publishing Corporation** (https://www.hindawi.com/spotlight/) – видавець журналів відкритого доступу, які пройшли експертну оцінку (містяться статті з різних галузей науки, техніки);

– **IntechOpen** (https://www.intechopen.com/) – видавець журналів та книг відкритого доступу з різних галузей науки, техніки (у видавництві публікуються понад 117 тисяч авторів з 154 країн світу);

– **J-STAGE** (https://www.jstage.jst.go.jp/browse/) містить близько 3000 наукових журналів (близько 5 млн. статей) з різних галузей знань;

– **PNAS** Proceedings of the National Academy of Sciences of the United States of America (https://www.pnas.org/) містить статті з фізики, біології, хімії, фармакології, математики;

– **Science.gov** (США) (https://www.science.gov/) здійснює пошук у понад 60 базах даних та понад 2200 наукових веб-сайтів і надає користувачам доступ до понад 200 мільйонів сторінок наукової інформації, включаючи результати досліджень та розробок;

– **WorldWideScience.org** (https://worldwidescience.org/indextext.html) – безоплатний електронний ресурс з доступом до національних та міжнародних науково-технічних баз даних та порталів з понад 70 країн світу;

– **ГПНТБ России** (Государственная публичная научно-техническая библиотека России) (http://www.gpntb.ru/) – каталог електронних ресурсів ДПНТБ Росії;

– **СО РАН** (http://sibran.ru/journals/) – посилання на журнали, що видаються СО РАН;

– **НАУКА** (https://naukapublishers.ru/) – електронна бібліотечна система, яка розроблена і підтримується видавництвом "Наука" (надає зручний віддалений доступ до повнотекстових архівів академічних журналів);

– **DOAJ** (Directory of Open Access Journals) (https://www.doaj.org/) – каталог електронних журналів відкритого доступу з різноманітної тематики, містить близько 14000 журналів із 130 країн світу;

– **SciELO** (The Scientific Electronic Library Online) (https://scielo.org/) – наукова електронна бібліотека, що охоплює колекцію наукових журналів з Бразилії та інших країн Латинської Америки і Карибського басейну, Португалії та Іспанії (містить майже 1250 журналів та понад 570 тис. статей);

– **UCSF Library (**University of California San Francisco) (https://www.library.ucsf.edu/journals/) – пошукова сторінка наукової бібліотеки університету Сан-Франциско, яка містить в своїй базі посилання на видання з різних напрямів, зокрема, 2556 онлайн-журналів з хімії;

– **eLIBRARY (**https://elibrary.ru/defaultx.asp) – наукова електронна бібліотека включає кілька тисяч повнотекстових наукових журналів в галузі природничих, гуманітарних наук, техніки, медицини і т. і. (для роботи з бібліотекою потрібна реєстрація, при цьому зміст журналів та реферати статей доступні безкоштовно для всіх зареєстрованих користувачів, а для працюючих з авторизованих IP-адрес конкретної організації – доступні повні тексти статей відповідно до договору);

– **Genamics JournalSeek** (http://journalseek.net/index.htm)– сайт, створений за підтримки OCLC, являє собою базу даних, яка міститиме відомості про майже 40 000 журналів (більше 6500 видавництв з посиланнями на їх сайти) з усіх галузей знань, розподілених за дворівневим рубрика тором (журнал в базі можна шукати за його назвою та або ISSN; у базі наводяться основні відомості про журнал і посилання на його адресу в мережі Інтернет);

– **Ingenta connect** https://www.ingentaconnect.com/?logoHome=true) надає алфавітний перелік журналів з можливістю перегляду змісту, а також бібліографічних описів (зокрема, рефератів) окремих статей (можливий перегляд журналів за видавництвами і темами; при цьому, інформація надається не з сервера видавництва, а з сервера компанії Ingenta);

– **Академия АНРИ** (https://academy.rasep.ru/listing) – навчально-консультаційний центр, який проваджує свою діяльність при Асоціації наукових редакторів і видавців **і безоплатно надає доступ до переліків журналів в індексах цитування Scopus і Web of Science (**доступні переліки різного ступеня актуальності).

Існують також спеціалізовані бази пошуку інформації про наукові періодичні видання світу з окремих галузей знань, зокрема, з хімії:

– **ABC-Chemistry** (http://abc-chemistry.org/) – каталог безкоштовних журналів з питань хімії;

– **CHEMNET** (http://www.chem.msu.su/) – офіційне електронне видання Хімічного факультету МДУ, яке містить посилання на різноманітні публікації з хімії;

– **ИКЦ «Академкнига» (**http://sciencejournals.ru/catalog/rubrics/27/) – наукові видання хімічного спрямування видавничо-книготоргового центру "Академкнига";

– **Chemistry Journals (**http://www-jmg.ch.cam.ac.uk/data/c2k/cj/) – один з найбільших списків посилань на журнали в галузі хімії, розміщений на сайті Department of Chemistry, University of Cambridge (Goodman Group) і організований за видавництвами, що видають журнали (можливий також перегляд в алфавітному порядку);

– **Organic Chemistry Journals** (https://www.organicdivision.org/journals/_ – сторінка, створена Division of Organic Chemistry Американського хімічного товариства, являє собою список журналів в галузі органічної хімії (надається доступ до змісту номерів журналів і список посилань на сайти видавництв, що видають ці журнали);

– **Журналы** – розділ порталу Аналітична хімія в Росії містить список посилань як на російськомовні, так і на іноземні журнали з хімії (близько 500 найменувань, http://www.wssanalytchem.org/resources/Lists/List/Journals.aspx) із зазначенням часового діапазону видання журналу і рівня надання безоплатної інформації.

Останнім часом до авторів наукових публікацій висувається вимога щодо публікації результатів досліджень в журналах, включених в глобальні індекси цитування **Scopus** і **Web of Science**, які являються одними з провідних і найкращих спеціалізованих систем пошуку наукової бібліографічної інформації.

Міністерство освіти і науки України забезпечило підключення понад 100 українських вишів, до яких входить і ДДМА, та наукових установ МОН до міжнародних наукових баз даних Scopus та Web of Science за кошти держбюджету. Доступ до цих баз можливий для користувачів внутрішньої мережі Академії.

Scopus (https://www.elsevier.com/solutions/scopus) – одна з найбільших уніфікованих реферативних баз даних рецензованої науково-дослідної літератури.

Власником та розповсюджувачем бази Scopus є компанія Elsevier – провідне видавництво, яке щороку випускає близько чверті всіх статей світових наукових журналів. У 1999 році видавництвом було створено платформу Sciencedirect – онлайнову базу даних ресурсів видавництва, яка стала важливим інформаційним ресурсом міжнародного науково-дослідного співтовариства. На основі Sciencedirect у 2002 році була створена база даних Scopus, яка охоплює науково-дослідну літературу з усього світу, а також якісні інтернет-джерела з ефективними інструментами для відстеження, аналізу та візуалізації досліджень.

Актуальну інформацію про видання, що реферуються базою даних Scopus, можна знайти за посиланням ***Download the Source title list*** на сторінці

https://www.elsevier.com/solutions/scopus/how-scopus-works/content. У файлі з розширенням *.XLSX, що надається системою, можна знайти ***українські наукові видання*** за допомогою фільтра “Ukraine” у стовпці X ("Publisher's Country/Territory"). На листопад 2019 року базою Scopus реферуються 48 наукових видань України.

Web of Science – база даних Інституту наукової інформації (https://clarivate.com/webofsciencegroup/solutions/web-of-science/), який було створено у 1960 році у Філадельфії, США (Institute for Scientific Information, ISI). Інститут наукової інформації займається складанням бібліографічних баз даних наукових публікацій, їх індексуванням і визначенням наукометричних показників. Основні продукти інституту:

- реферативні бази даних;
- щорічний звіт Journal Scitation Report, у якому наводяться імпакт-фактори всіх журналів, що індексуються інститутом;
- щорічний список найбільше цитованих вчених, на основі якого, зокрема, складається Академічний рейтинг університетів світу.

Інформаційну основу індексу цитування Web of Science складають індекси цитування з різних галузей наук, що відображені у базах даних Інституту наукової інформації, а також спеціалізовані покажчики, які об’єднують матеріали конференцій та симпозіумів, оглядові видання та ін.

Science Citation Index (SCI) та її інтернет-версія ***Web of Sciences*** (WOS) – база даних індексів цитування з природничих і точних наук, яка існує з 1963 року і охоплює публікації, що стосуються наук про життя і Землю, різних галузей фізики, хімії, математики, агронауки та тваринництва, харчової промисловості, сільського господарства, охорони навколишнього середовища, клінічної медицини, техніки і технології, прикладних наук і будівництва. Це – спеціалізований інформаційний продукт, в якому збирається та обробляється бібліографічна інформація про наукові публікації, а саме:

- назва тексту;
- вихідні дані;
- автор тексту (прізвище, ім’я, назва організації, в якій він працює);
- тип тексту (стаття, доповідь, рецензія тощо);
- ключові слова;
- список цитованої літератури;
- мова, якою опубліковано текст.

База дозволяє знаходити як публікації, що цитуються у кожній окремо взятій статті, так і публікації, що цитують цю статтю. Таким чином, бази даних ISI можуть використовуватися як механізм, що дає можливість інтегрувати публікаційні та цитатні показники по всій вертикалі соціального інституту на рівні від наукового співробітника – автора, структурного підрозділу і наукової організації, де працюють більшість авторів, до міністерств і відомств або цілих адміністративно-географічних регіонів.

Google Scholar (https://scholar.google.com/) – це пошукова система вільного доступу, яка забезпечує повнотекстовий пошук наукових публікацій усіх форматів та дисциплін.

Система працює з листопада 2004 року. Індекс Академії Google включає в себе більшість рецензованих онлайн-журналів Європи та Америки найбільших наукових видавництв.

Академія Google дозволяє користувачам здійснювати пошук цифрової або фізичної копії статей, онлайн або в бібліотеках. Наукові результати пошуку генеруються з використанням посилань з повнотекстових журнальних статей, технічних звітів, препринтів, дисертацій, книг та інших документів, у тому числі обраних веб-сторінок, які вважаються науковими. Оскільки більшість наукових результатів пошуку Google – це прямі посилання на комерційні журнальні статті, більшість користувачів зможуть лише отримати доступ до анотації статті, а також невеликої кількості важливої інформації про статтю, можливо, доведеться платити за доступ до повного тексту.

Index Copernicus (IC) (https://journals.indexcopernicus.com/) – міжнародна платформа для просування спеціалізованих наукових досягнень. Вона також сприяє підтримці національного та міжнародного співробітництва між вченими й видавцями наукових журналів. Платформа IC була створена у Польщі у 1999 році. Щорічно вона публікує рейтинг видань та проводить детальний аналіз не тільки наукового рівня кожного журналу, але і його оформлення, розповсюдженості, привабливості для читачів тощо.

Корисні ресурси для знаходження наукових публікацій — загальнонаукові пошукові системи видавців академічних публікацій. Найбільші такі ресурси – це **Science Direct** (https://www.sciencedirect.com/) і портал журналів Американського хімічного товариства **ACS Publications** (https://pubs.acs.org/).

Багато профільних журналів доступні також на сайтах найбільших видавців: **Springer** (https://www.springer.com/gp), **Wiley Online Library** (https://onlinelibrary.wiley.com/), **Oxford University Press Academic** (https://global.oup.com/academic/?cc=ua&lang=en&) та ін.

Американський центр National Center for Biotechnology Information, що входить у систему Національних Інститутів Здоров'я (NIH), створив ряд баз даних вільного доступу високого рівня. Надзвичайно популярною у фахівців у галузі наук про життя є система **PubMed.gov** (https://pubmed.ncbi.nlm.nih.gov/). Це бібліографічна база даних, що надає можливість здійснювати ефективний пошук наукової літератури за ключовими словами.

Безумовно, слід мати на увазі, що в більшості випадків вебресурси (наукові статті, інформація з баз даних тощо), знайдені науковими пошуковими системами, є комерційними й недоступними в повному обсязі. Сьогодні це є однією з проблем Інтернету, оскільки достатньо легко знайти потрібні дані, але вони не перебувають у вільному доступі. Слід зазначити, що абстракти статей практично завжди відкриті, а доступ до повнотекстових публікацій найчастіше є платним. Видавці деяких журналів пішли іншим шляхом – їх повнотекстові публікації знаходяться у вільному доступі, проте авторам статті доводиться сплачувати вартість її публікації.

Багато інформації можна відшукати і в наукових журналах із відкритим доступом (***Open Access***). Велику кількість таких видань розміщено на порталі BioMed Central (https://www.biomedcentral.com/).

2.2.3 Патентні бази даних

Патентна інформація стає все більш актуальною для хіміків-синтетиків, причому повнотекстові патенти, в цілому, значно доступніші за статті в академічних журналах.

Пошук патентів США доступний через веб-сайт **USPTO** за адресою: https://www.uspto.gov/patents-application-process/search-patents. Патенти у форматі HTML доступні з 1976 року. Починаючи з 1790 року, вони представлені у вигляді посторінково відсканованих документів (TIF файли).

Пошук патентів можливий також через пошукову систему **Google Patents** (https://patents.google.com/).

Номер патенту є «магічним» ключем до патентної пошукової інформаційної системи. Незалежно від того, в яку дату був виданий патент, якщо відомий номер патенту, можна швидко отримати повнотекстовий патент, використовуючи безкоштовні інструменти, доступні в Інтернеті. Практично всі веб-сайти для безкоштовного пошуку патентів дозволять ввести номер патенту та отримати його PDF-версію.

Якщо відомий номер патенту, для пошуку можна використовувати як ***Google Patents*** так і веб-сайт ***USPTO*** – необхідно тільки ввести номер патенту без коми у пошукову строку. При використанні веб-сайту USPTO патент повинен бути довжиною сім номерів (якщо потрібно, додайте попередні нулі).

Якщо номер патенту невідомий, можна шукати патенти за темою чи винахідником, але варіанти залежать від того, наскільки старі патенти необхідні.

Патенти, видані після 1975 року, можна легко шукати за ключовими словами, використовуючи ряд безкоштовних інструментів.

Для патентів, виданих у 1975 році або раніше варіанти пошуку є складнішими. Багато ранніх патентів зараз можна шукати в повнотекстовому форматі за допомогою ***Google Patents***, хоча цифрові тексти створюються автоматично і можуть містити помилки або бути важкими для пошуку.

Патенти США, видані з 1790 по 1975 роки, можна шукати на веб-сайті ***USPTO*** лише за датою випуску, номером патенту або класифікаційним кодом.

Екран розширеного пошуку (https://patents.google.com/advanced) патентів Google дозволить здійснити пошук за номером патенту, винахідником, правонаступником, темою, номером класифікації та датою. Після пошуку можна уточнити результати за датою випуску, датою публікації, патентним відомством, статусом подачі та типом патенту.

Більшість інших основних баз патентів також є безплатними. Серед них

– **EPO Espacenet** (https://worldwide.espacenet.com/), асоційований з Європейським патентним офісом;

– **WIPO IP Portal** (https://patentscope.wipo.int/beta/ru/search.jsf) – пошук по національним патентним фондам і фондам PCT (Patent Cooperation Treaty);

– **GOV.UK** (https://www.gov.uk/topic/intellectual-property/patents) – база Британського патентного офісу;

– **CANZLER & BERGMEIER** (https://www.cb-patent.com/en/ip-rights/patent/) – база європейських патентів;

– **FPO** (http://www.freepatentsonline.com/search.html) – пошук патентів по національним патентним фондам (US Patents, US Patent Applications, EP documents, Abstracts of Japan, WIPO (PCT), German Patents);

– **Федеральный Институт Промышленной Собственности** (https://www1.fips.ru/elektronnye-servisy/) – широка база патентів РФ, авторських свідоцтв та патентів колишнього СРСР (містить зручну пошукову систему і широку базу відкритих реєстрів);

– **Derwent World Patents Index (DWPI)** https://clarivate.com/derwent/solutions/derwent-world-patent-index-dwpi/ – найповніша в світі добірка патентних документів з професійними анотаціями і коментарями, отриманих від 40 патентних бюро в усьому світі. Всі дані патентів регулярно оновлюються новою інформацією, включаючи дані, пов'язані з Ресурсом з хімії Derwent, унікальною базою даних структурних формул;

– **УКРПАТЕНТ** (https://ukrpatent.org/uk) – державне підприємство "Український інститут інтелектуальної власності" (Укрпатент) – інституційна складова державної системи правової охорони інтелектуальної власності в Україні; пошук патентів на винаходи і патентів на корисні моделі краще здійснювати на сторінці https://ukrpatent.org/uk/articles/bases2 за посиланням ***Спеціалізована БД "Винаходи (корисні моделі) в Україні"***.

У всіх патентних базах можливий пошук як за номером патенту, так і за ключовими словами або назвою патенту.

2.2.4 Системи ідентифікації науковців і оцінювання наукової діяльності

Проблема точної ідентифікації автора наукової роботи не є новою у світі. Більшість імен протягом життя можуть змінюватися, хибно транслітеруватися. Щодня свій науковий доробок публікують у різноманітних наукових виданнях десятки науковців з однаковими прізвищами у різних країнах. До того ж, у міжнародних журналах наші автори іноді публікуються під різними варіантами свого прізвища (наприклад, Олійник – Oliynyk, Oliinyk, Oliynik). Через це при підрахунку цитувань робіт вчених міжнародними наукометричними базами виникає чимало плутанини та помилок. Вирішенням проблеми є використання унікального ідентифікатора авторів-науковців – ID (unique author identifier).

ID науковця дозволяє:

–легко встановити, хто є автором конкретного документу;

–точно виміряти цитованість робіт окремих дослідників;

–полегшує процес оцінки продуктивності та впливовості конкретного автора;

–спрощує обробку та зберігання даних в одному місці;

–покращує видимість публікацій автора у глобальній мережі.

Існують різні системи унікальних ідентифікаторів науковців: міжнародні та національні, мультидисциплінарні та галузеві.

ResearcherID – міжнародна ідентифікаційна система, що дозволяє створити унікальний профіль дослідника, який містить відомості про його наукові публікації і їх історії. В даний час є власністю Clarivate Analytics. Має наступний загальний вид: A-1234-2015, де A – літера латинського алфавіту, 1234 – значення від 001, до 9999, 2015 – рік реєстрації ідентифікатора.

Система була створена і введена в дію в січні 2008 року компанією Thomson Reuters. Вона забезпечує обмін даними між своєю базою даних і базою даних ORCID. Однак, суть цих двох ідентифікаторів, незважаючи на однакові цілі, різниться. ORCID не пов'язаний з якою-небудь базою даних в той час, як ResearcherID є інструментом, тісно пов'язаним з базою Web of Science, що значно розширює його можливості і дещо спрощує створення і підтримання профілю.

З 2016 року система підтримується компанією Clarivate Analytics. З 2019 року колишній ресурс researcherid.com скасований і розпочато перенесення інструментів на платформу **Publons**, яка надає інструменти рецензування наукових публікацій. Видача ResearcherID тепер можлива лише при наявності в профілі однієї або більше власних публікацій. Порожні профілі не можуть отримати ID. Авторам, що вже мали на момент переїзду на платформу Publons ідентифікатори, були розіслані запрошення. Зареєструватися в **Publons** можна за посиланням https://publons.com/in/researcherid/.

Профілі **ResearcherID** і **Publons** об'єдналися, щоб використовувати всі можливості систем ***Web of Science***, ***ResearcherID*** і ***Publons*** за допомогою єдиного облікового запису.

Профіль **Publons** надає користувачеві наступні можливості:

- всі публікації автора легко імпортуються з Web of Science, ORCID або з іншого менеджера бібліографічних посилань (наприклад, EndNote, Zotero або Mendeley);
- надійні показники цитування, які автоматично оновлюються на основі даних Web of Science Core Collection, що містить понад 21 000 кращих світових журналів;
- перевірену історію рецензування і редагувань, засновану на партнерстві з тисячами наукових журналів;
- завантаження академічного звіту, в якому коротко описується вага науковця як автора, редактора та рецензента, який він можете використовувати для просування по службі та кар'єрних переговорів.

ORCID (https://orcid.org/) – це відкритий, некомерційний проект для створення і підтримки реєстру унікальних ідентифікаторів дослідників, прозорого способу ув'язки науково-дослідної діяльності та доступу до цих іден-

тифікаторів.

ORCID є унікальним завдяки своїй незалежності від наукових дисциплін і національних кордонів, а також взаємодії з іншими системами ідентифікації. Основною метою введення системи ORCID є можливість ідентифікації наукових робіт, написаних різними вченими з однаковими іменами та прізвищами. Ідентифікатор представляє собою 16-значне число, унікальне для кожного автора.

Обліковий запис ORCID включає в себе інформацію про ім'я вченого, його електронну адресу, назву організації та його дослідницьку діяльність. ORCID враховує необхідність контролю за поширенням цих даних і надає відповідні інструменти для управління рівнем приватності даних.

Структура ідентифікатора ORCID ID являє собою номер з 16 цифр, узгоджений із стандартом ISO (ISO 27729). Крім цифр від 0 до 9 ідентифікатор може містити велику літеру X, що представляє число 10. ORCID ID відображається як адреса виду http://orcid.org/xxxx-xxxx-xxxx-xxxx.

На сьогодні членами ORCID є близько 300 організацій, зокрема, чимало авторитетних університетів та наукових видавництв. ***ORCID*** – одна з небагатьох систем, що ***дозволяє пов'язати різні унікальні ідентифікатори автора***. Це є важливим, враховуючи кількість систем ідентифікації й те, що автор може бути зареєстрований у декількох з них.

Scopus Author ID – для авторів, які опублікували більше однієї статті, у ***Scopus*** (https://www.elsevier.com/solutions/scopus) створюються індивідуальні облікові записи – профілі авторів з унікальними ідентифікаторами авторів (**Author ID**).

Ці профілі надають таку інформацію, як варіанти імені автора, перелік місць його роботи, кількість публікацій, роки публікаційної активності, галузі досліджень, посилання на основних співавторів, загальна кількість цитувань на публікації автора, загальна кількість джерел, на які посилається автор, індекс Хірша автора тощо. База даних надає користувачам можливості використання унікальних ідентифікаторів авторів для формування пошукових запитів та налаштування сповіщень (електронною поштою або через RSS) щодо змін у профілях авторів. Можливості пошуку авторів та обмеженого перегляду їх профілів доступні без наявності передплати на базу даних Scopus засобами ***Scopus Author Preview*** (https://www.scopus.com/freelookup/form/author.uri).

Для установ, співробітники яких опублікували більше однієї статті, в Scopus створюються профілі з унікальними ідентифікаторами установ – Scopus Affiliation Identifier. Ці профілі надають таку інформацію, як адреса установи, кількість авторів-співробітників установи, кількість публікацій співробітників, перелік основних назв видань, в яких публікуються співробітники установи, і діаграма тематичного розподілу публікацій співробітників установи.

База даних Scopus надає широкі можливості отримання наукометрії і проведення автоматизованого аналізу видань. Інструмент Journal Analyzer дозволяє проводити розширений аналіз наукового рівня видань (в тому числі, порівняльний аналіз декількох видань) за чотирма основними показниками:

–загальне число статей, опублікованих у виданні протягом року;

–загальна кількість посилань на видання в інших виданнях протягом року;

–тренд року (відношення кількості посилань на видання до кількості статей, опублікованих у виданні);

–відсоток статей, що не були процитовані.

База даних Scopus в багатьох країнах є одним з головних джерел отримання наукометричних даних для проведення оціночних досліджень на державному та/або корпоративному рівні.

Дані Scopus спочатку використовувалися в рейтингу провідних університетів світу Times Higher Education Supplement: World University Rankings (QS TopUniversities). Однак, в останні роки цей аналіз і рейтинг проводиться на основі бази даних Web of Science компанії Clarivate Analytics.

Індекс Хірша (h-індекс) – наукометричний показник, запропонований в 2005 році аргентино-американським фізиком Хорхе Хіршем з Каліфорнійського університету в Сан-Дієго.

Індекс Хірша пропонувався як ***кількісна характеристика продуктивності вченого, групи вчених, наукової організації або наукової спільноти країни*** в цілому, що ***оцінюється за кількістю публікацій і цитувань цих публікацій***.

Індекс Хірша був розроблений як альтернатива класичним «індексам цитованості» – сумарному числу посилань на роботи вченого. Критерій засновано на сукупному обліку числа публікацій дослідника і числа цитувань цих публікацій. Вчений має індекс h, якщо h з його N статей цитуються як мінімум h раз кожна.

Наприклад, h-індекс дорівнює 10. Це означає, що вченим було опубліковано щонайменше 10 робіт, кожна з яких була процитована 10 і більше разів. При цьому кількість робіт, процитованих менше число раз, може бути будь-яким, і воно не дає вкладу в індекс Хірша. Таким чином, для досягнення високого індексу Хірша недостатньо мати багато публікацій і навіть високий індекс цитованості, а важливо, щоб частіше цитувалася як можна більша кількість опублікованих робіт. Тобто, h-індекс – це спроба дати комплексну оцінку одночасно числу публікацій вченого і їх цитованості (якості). Індекс Хірша був придуманий, як уніфікована оцінка ефективності праці вченого незалежно від сфери його досліджень.

Індекс Хірша може обчислюватися з використанням як безкоштовних загальнодоступних наукометричних баз даних в Інтернеті, так і баз даних з платної підпискою (наприклад, Scopus або ISI Web of Science). Однак платні бази даних часто теж розміщують h-індекс вчених у вільному доступі. Індекс Хірша, підрахований для одного і того ж науковця з використанням різних баз даних, буде різний, як і інші наукометричні характеристики. Він залежить від кількості наукових публікацій автора в виданнях, що увійшли до обраної бази даних. Крім того, індекс Хірша може обчислюватися з урахуванням і без урахування самоцитування; передбачається, що відкидання посилань авторів на власні статті дає більш об'єктивні результати. Наприклад, в рейтингу вчених

України за індексом Хірша виконується підрахунок за даними бази Scopus з відкиданням самоцитування всіх авторів (тобто цитування статті 1 в статті 2 цієї статті не враховується, якщо хоча б один автор входить одночасно в список співавторів обох статей).

Щоб підвищити індекс Хірша необхідно:

– публікувати оригінальні статті високої наукової значущості;

– намагатися публікувати як свої статті, так і наукові статті в співавторстві з авторитетними вченими;

– публікувати статті не тільки в національних фахових журналах, а й в англомовних виданнях , що входять до міжнародних баз даних (Scopus, Web of Science).

– обмінюватися посиланнями зі своїми колегами;

– обов'язково перевіряти правильність написання особистих даних при подачі заявки на публікацію статті;

– уважно стежити за індексацією вашої статті в різних базах.

Російський індекс наукового цитування (**РИНЦ**) – бібліографічна база даних наукових публікацій російських вчених. Для отримання необхідних користувачу наукометричних даних про публікації та цитованість статей на основі бази даних РИНЦ розроблено аналітичний інструментарій ***ScienceIndex***. Проект РИНЦ розробляється з 2005 року компанією «Наукова електронна бібліотека» (https://elibrary.ru/defaultx.asp).

DOI (digital object identifier) https://www.doi.org/ – ідентифікатор цифрового об'єкта, який присвоюється науковим статтям і збірникам.

У DOI може входити різна інформація, наприклад, адреса статті в Інтернеті (URL – Uniform Resource Locator), назва статті, автор, інформація про видання та інші метадані, які Ви повинні ретельно заповнити і перевірити ще раз при додаванні статей і збірок в Open Journal Systems.

С ідентифікаторами цифрових об'єктів (DOI) наукові журнали мають набагато більше шансів бути включеними та індексуватися в авторитетних міжнародних наукометричних базах даних Scopus і Web of Science.

2.2.5 Основи роботи в міжнародній науковій базі Scopus

МОН України забезпечило підключення понад 100 українських вишів, зокрема, і ДДМА, до міжнародних наукових баз даних Scopus та Web of Science за кошти держбюджету.

Доступ до міжнародної наукової бази даних Scopus можливий для користувачів внутрішньої мережі Академії за адресою – https://www.scopus.com/search/form.uri?display=basic. Система автоматично відкриває вкладку пошуку ***Document search (Поиск документов)*** (рис. 2.19).

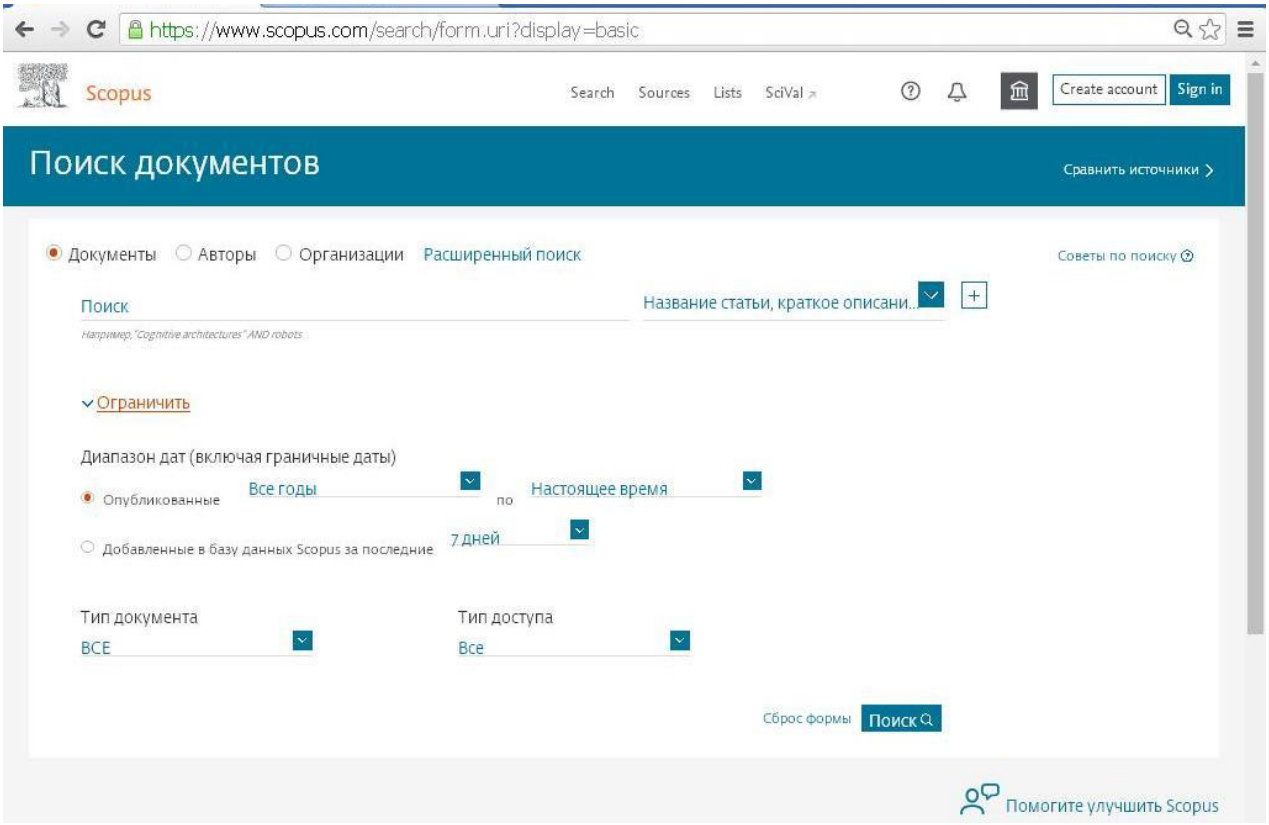


Рисунок 2.19 – Робоче вікно пошуку документів бази даних Scopus

Scopus пропонує кілька видів пошуку:

- пошук документів;
- пошук авторів;
- пошук організації;
- розширений пошук.

При пошуку документів можна задати декілька ключових запитів, розділивши їх оператором AND (рис. 2.20).

Можна також вибрати критерії пошуку (див. рис. 2.20):

- ***Все поля***;
- ***Название статьи,краткое описание, ключевые слова;***
- ***Авторы;***
- ***Первый автор*** и т. і.

Можна ввести обмеження щодо:

– року видання;

– строку додавання запису в Scopus;

– типу документу (наприклад, ***все***; ***статья***; ***статья или обзор***, ***обзор***, ***книга*** і т. і.).

Пошук здійснюється натисканням кнопки ***Поиск*** (див. рис. 2.20).

На сторінці результатів пошуку (рис. 2.21) в лівій колонці можна уточнити результати – обмежити або виключити певні дані за наступними параметрами: ***тип доступу***, ***рік***, ***автор***, ***галузь знань***, ***тип документу*** та ін.

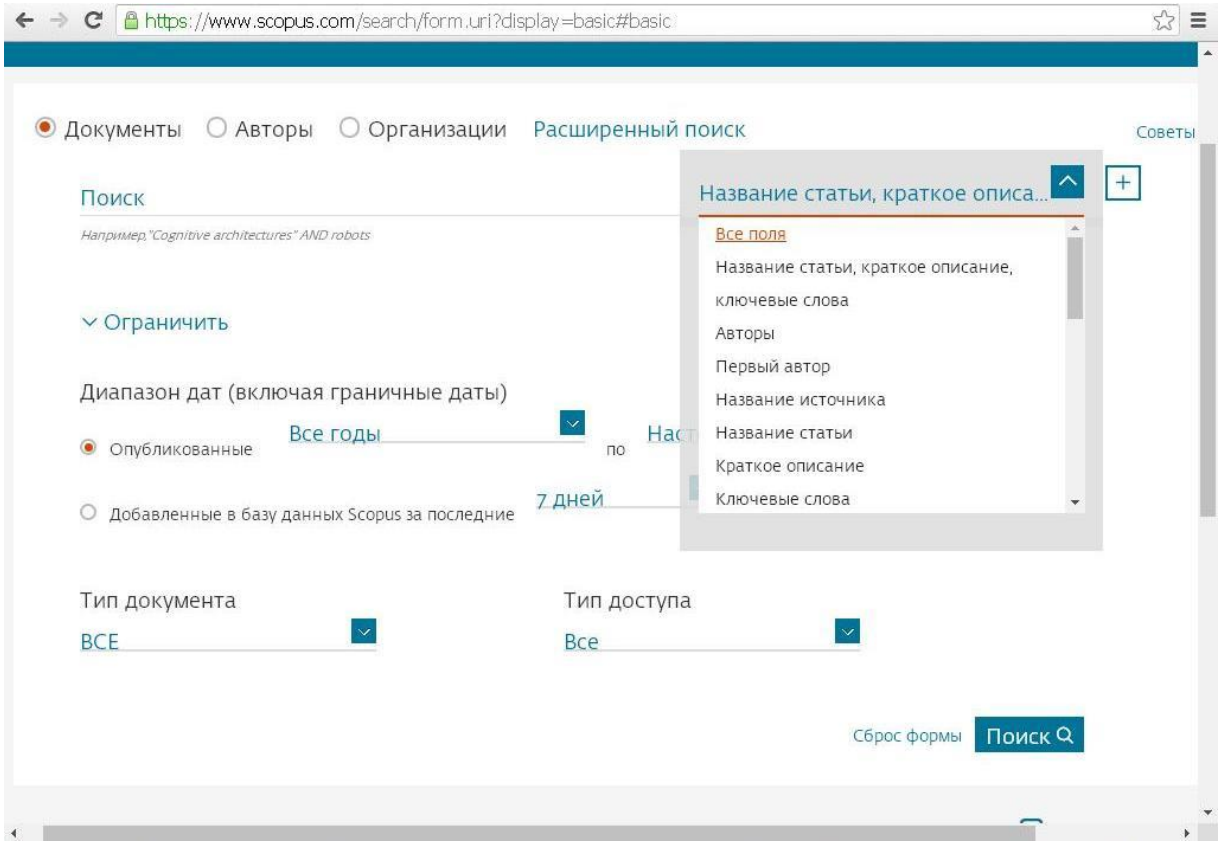


Рисунок 2.20 – Вікно пошуку документів бази даних Scopus

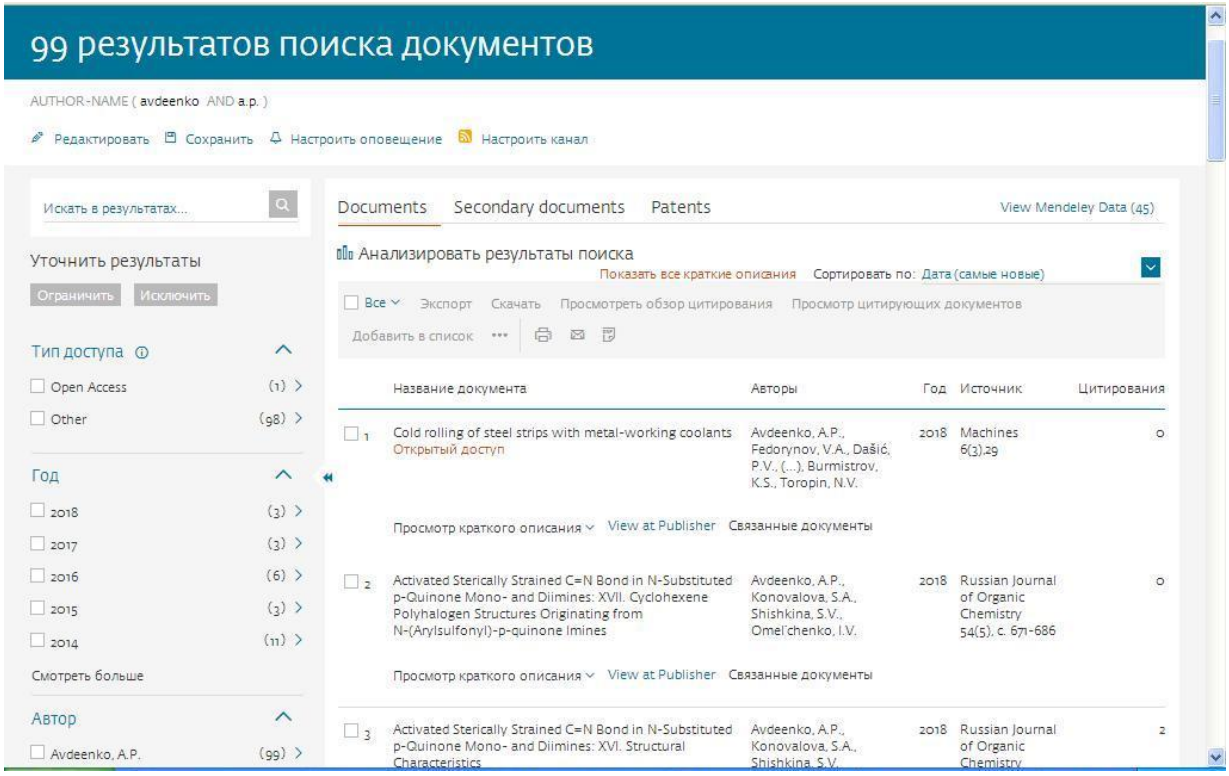


Рисунок 2.21 – Сторінка результатів пошуку документів бази даних Scopus

Сторінка результатів пошуку містить декілька вкладок (рис. 2.21):

- ***Documents***;
- ***Secondary documents***;
- ***Patents***.

На вкладці ***Documents*** наведено перелік документів (статей), які відповідають умовам пошуку. Статті можна сортувати за роком видання, за автором, за кількістю цитувань, за назвою джерела (журналу). З кожним документом (статтею) можна ознайомитися окремо – для цього треба вибрати відповідну дію (див. рис. 2.21):

- переглянути дані про документ, які містяться у базі Scopus (за натисканням на назву статті);
- переглянути анотацію (***Просмотр краткого описания***);
- переглянути статтю на сайті видавника (***View at Publisher***);
- переглянути перелік статей, що мають однакові посилання (***Связанные документы***);
- переглянути перелік цитувань (за натисканням на кількість цитувань);
- переглянути відомості про кожного автора (за натисканням на фамілію автора) (рис. 2.22);
- переглянути відомості про джерело/журнал (за натисканням на назву джерела) (рис. 2.23).

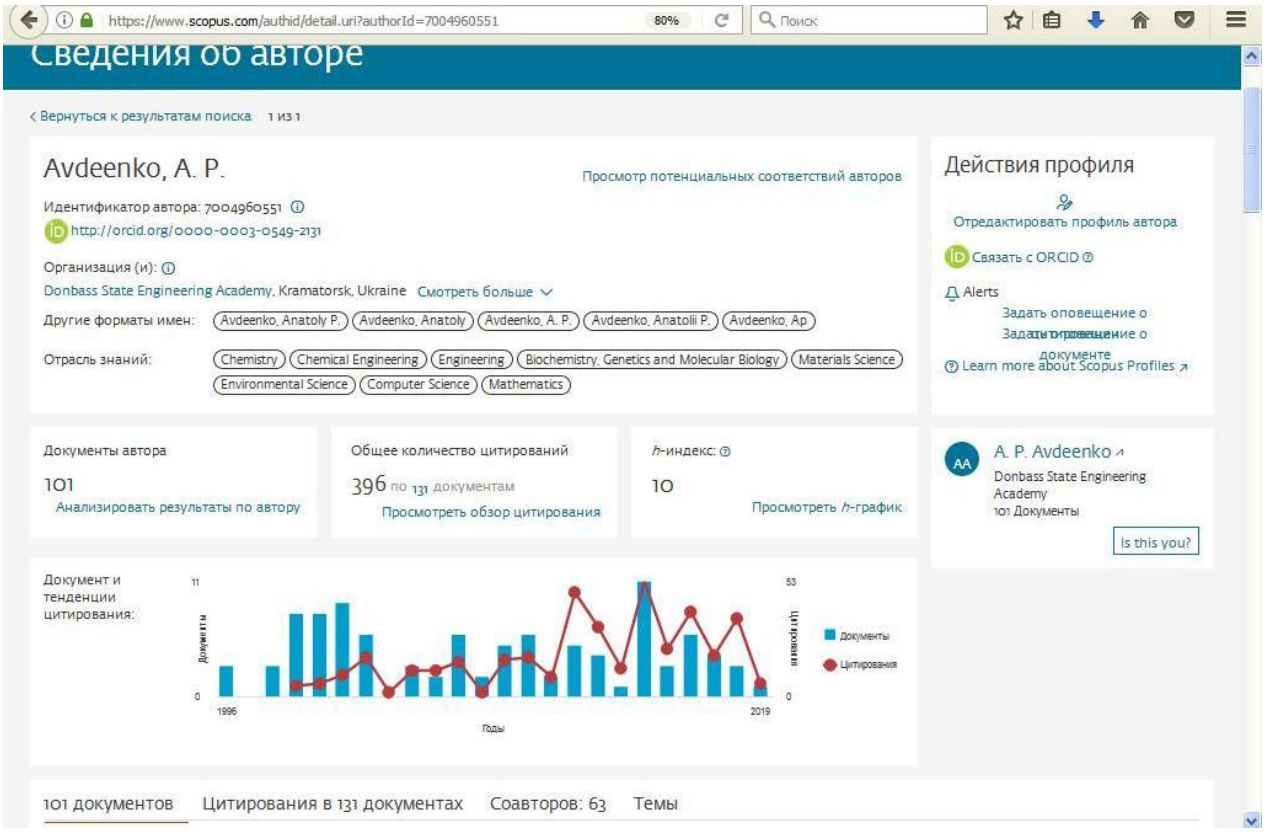


Рисунок 2.22 – Відомості про автора, що містяться у базі даних Scopus

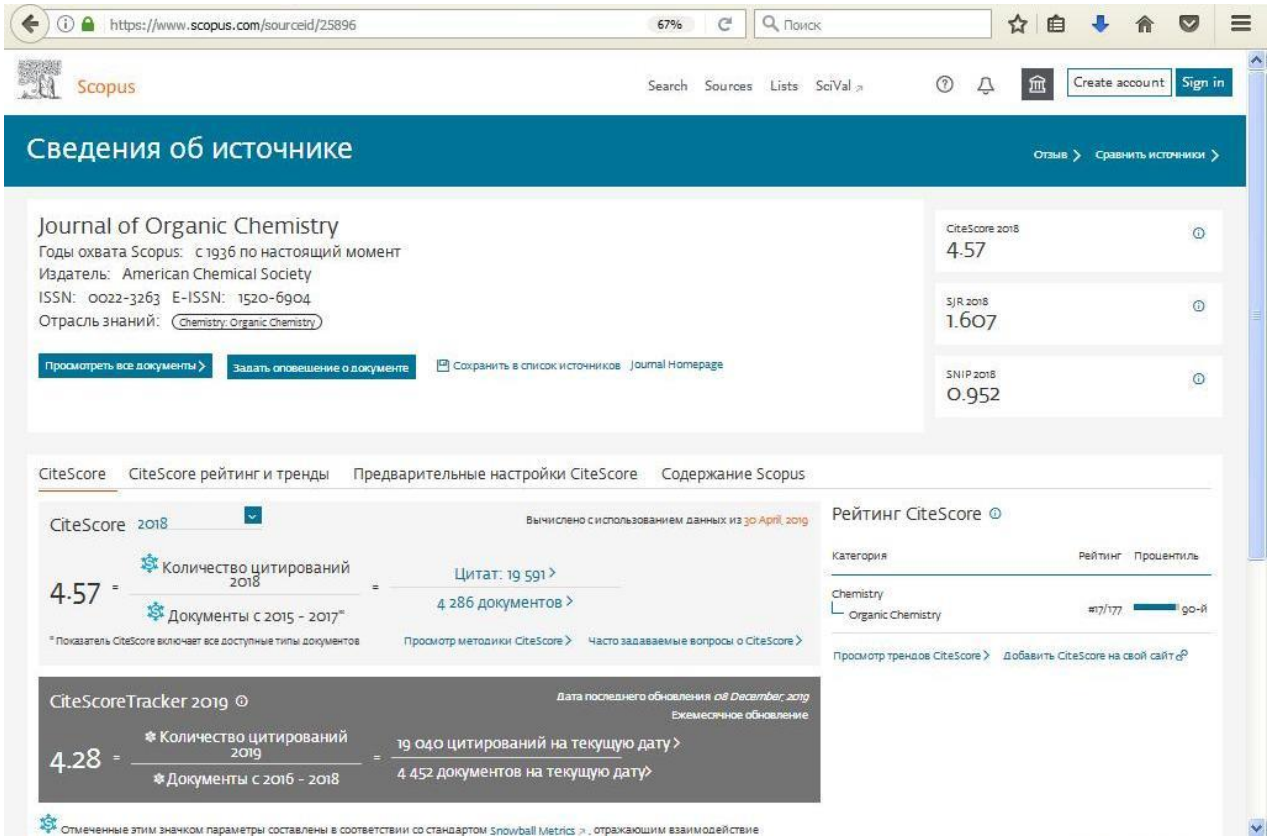


Рисунок 2.23 – Відомості про джерело (журнал), що містяться у базі даних Scopus

Вкладка ***Secondary documents*** містить перелік вторинних документів. Це документи, які були отримані із довідкового списку документів Scopus, який недоступний безпосередньо в цієї базі, оскільки він не індексується базою Scopus.

Вкладка ***Patents*** містить перелік патентів, які відповідають умовам пошуку.

На сторінці результатів пошуку також є посилання ***View Mendeley Data***, яке направляє до сторінці сайту ***Mendeley***, яка містить перелік експериментальних даних, зокрема, даних рентгеноструктурного аналізу відповідного документу (статті) (рис. 2.24).

При натисканні на назву відповідних даних сайт ***Mendeley*** перенаправляє на офіційний сайт джерела даних, зокрема, на сайт CCDC, який надає дані рентгеноструктурного аналізу (рис. 2.25).

При пошуку автора у Scopus необхідно перейдіть на вкладку ***Поиск авторов***. Далі треба ввести прізвище автора, ініціали, організацію та/або ORCID. Для пошуку необхідно натиснути ***Поиск*** (рис. 2.26).

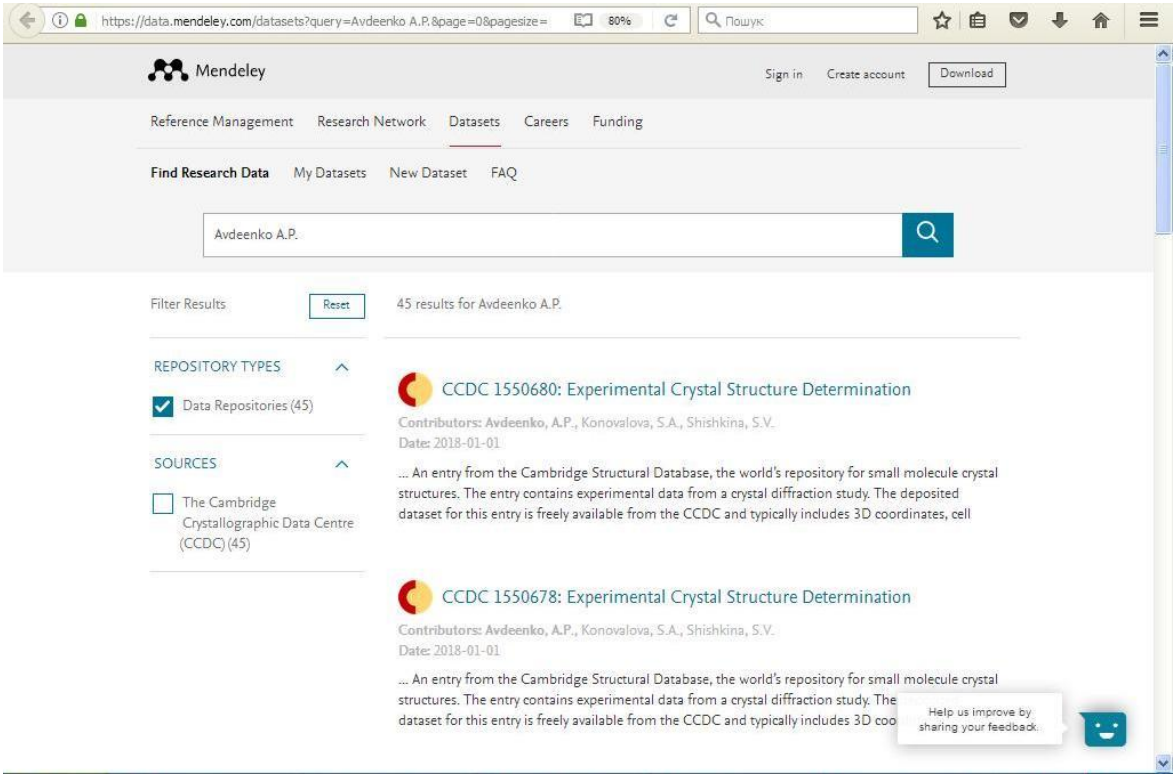


Рисунок 2.24 – Перелік даних рентгеноструктурного аналізу сайту Mendeley

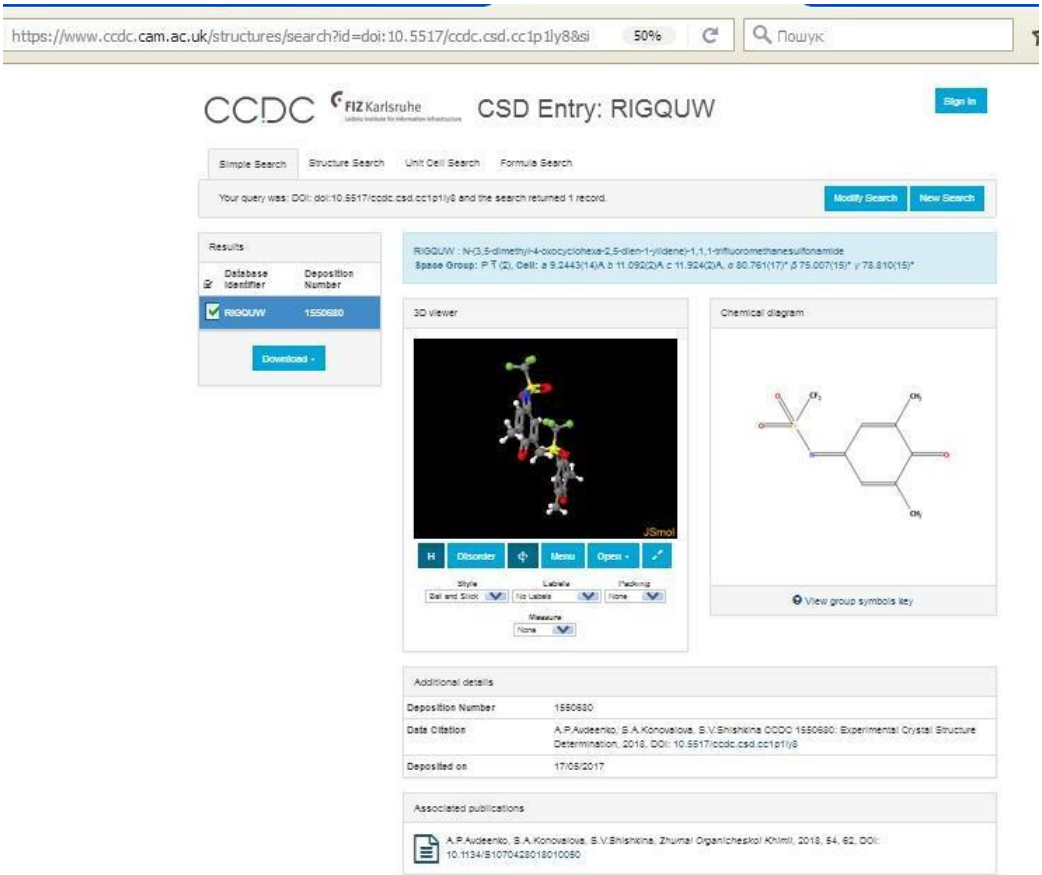


Рисунок 2.25 – Дані рентгеноструктурного аналізу

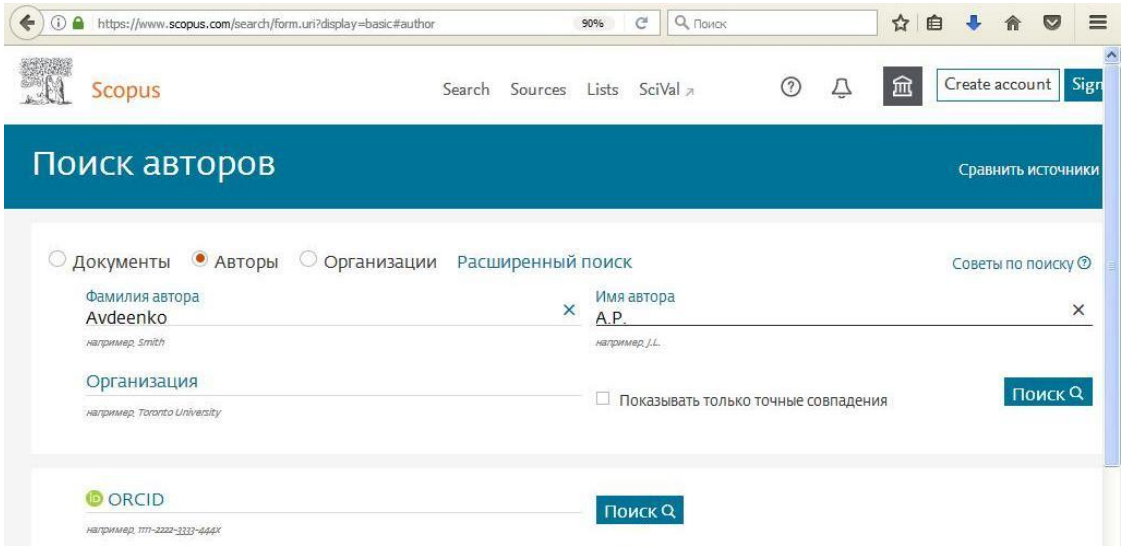


Рисунок 2.26 – Вікно пошуку авторів бази даних Scopus

Пошук авторів допомагає знайти в Scopus документи, написані певною людиною, навіть якщо ім'я автора наведено інакше. Наприклад, в одному документі ім'я автора може бути записано як Smith, J, а в іншому – Smith, John. ***Ідентифікатор учасника Scopus*** забезпечує ідентифікацію при пошуку у разі різних найменуваннях автора.

Функція «Ідентифікатор авторів Scopus» допомагає розрізняти авторів з поширеними іменами, наприклад Smith або Lee, пропонуючи список можливих відповідних авторів із зазначенням їх організацій та галузей знань, в яких вони працювали.

Наприклад, якщо в результаті пошуку знайдені два учасника з ім'ям Ayre, G., можна переглянути їх організації або галузі знань, в яких вони працювали, щоб вибрати для пошуку правильного автора з ім'ям Ayre, G.

Функція «Ідентифікатор учасника Scopus» також допомагає знаходити авторів, які по-різному цитувалися. Групуючи імена авторів під одним унікальним ідентифікаційним номером учасника, Scopus враховує різне написання прізвища автора, всі можливі поєднання імені та прізвища автора, а також написання прізвища автора з ініціалами і без ініціалів.

В результаті в умови пошуку за конкретним автором включаються переважне ім'я автора і різні варіанти написання цього переважного імені.

Наприклад, при пошуку по автору Ayre, G в результати включаються документи, ім'я автора яких написано як Ayre, Gareth і Ayre, G.

Також можна шукати автора за номером ORCID (Open Researcher and Contributor ID – відкритий ідентифікатор дослідників і учасників) (див. рис. 2.26), що значно спрощує процедуру пошуку.

В результаті пошуку, крім загальних даних про автора (див. рис. 2.23), система Scopus надає повний перелік публікацій даного автора (рис. 2.27).

Система Scopus надає також можливість знайти дані про певне видання. При цьому можна уточнювати критерії пошуку – вибрати відповідну галузь або науковий напрям, режим доступу і т.і. (рис. 2.28).

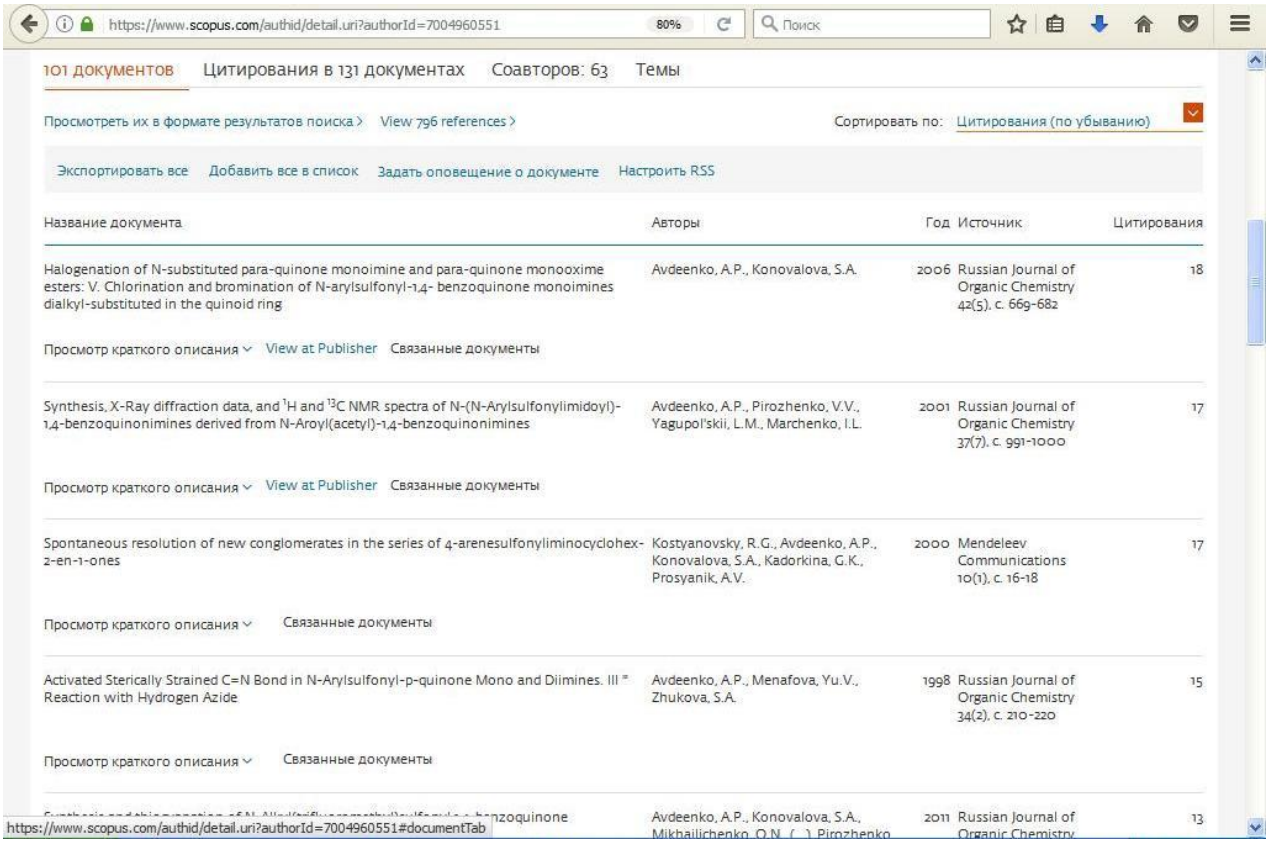


Рисунок 2.27 – Перелік публікацій певного автора, що містяться у базі даних Scopus

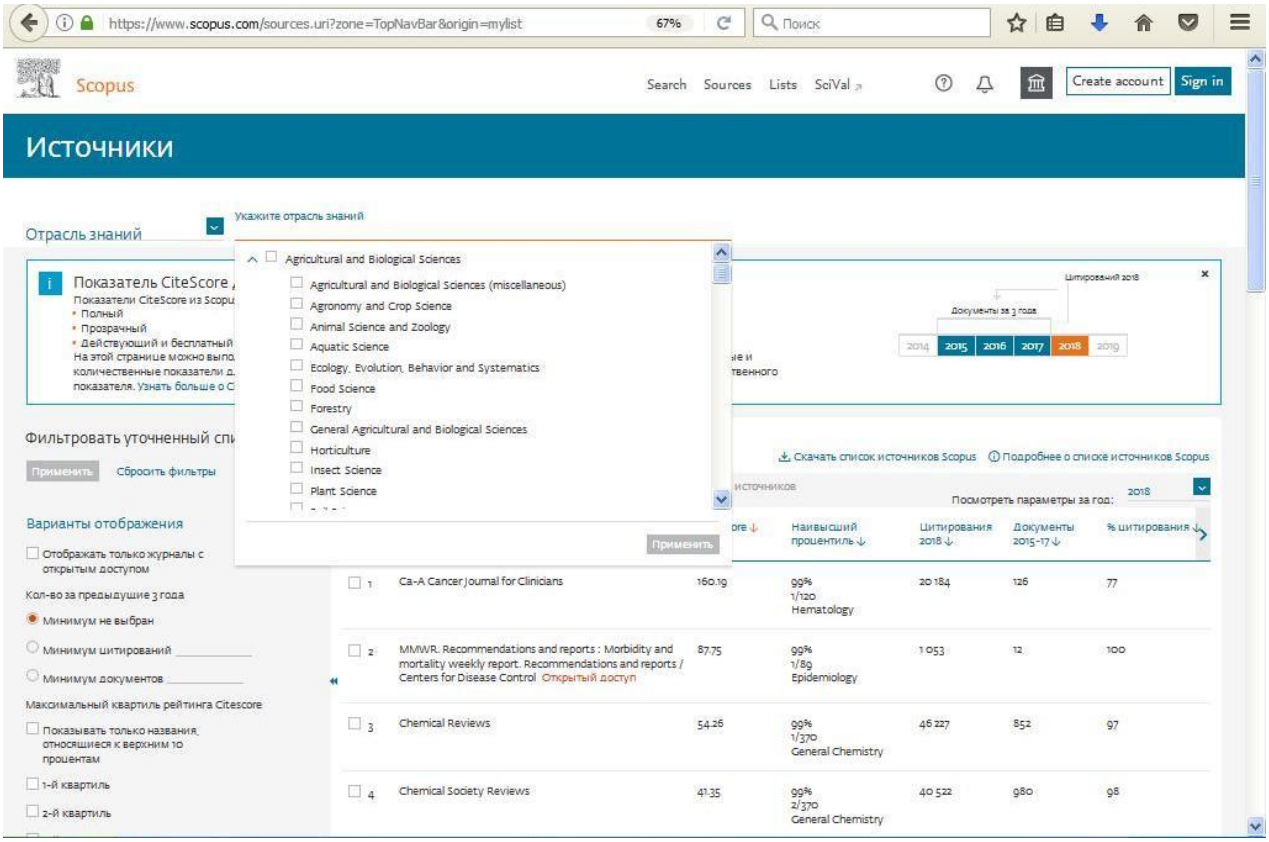


Рисунок 2.28 – Вікно пошуку видань (джерел) у базі даних Scopus

2.2.6 Пошук інформації з хімії у мережі Internet

За даними фахівців IBM кількість інформації на планеті подвоюється кожні 1,5–2 роки. Кожні 2,5–3 роки подвоюється число заявок на наукові відкриття та винаходи. Таке експоненціальне зростання інформації призводить до того, що людині, вченому, який працює навіть у вузькій області, стає все важче відстежувати результати своїх колег, а значить зростають нераціональні витрати часу і коштів на «відкриття велосипеда», повторення того, що вже зроблено іншими, опубліковано, запатентовано.

Для того, щоб швидше й ефективніше отримати потрібну інформацію, в Інтернеті існують ***шляхи оптимізації пошуку***. Насамперед необхідно максимально коректно задавати умови пошуку для мінімізації кількості отриманих посилань (часто їх сотні тисяч і мільйони).

Одним із найпростіших методів є пошук фрази (як правило, її треба взяти в лапки) замість набору ключових слів. Так, на запит ***quinoneimine synthesis*** пошукова система Google видає понад 90 000 посилань, тоді як на запит у форматі «***quinoneimine synthesis***» – лише 3.

Інший шлях – ***використання логічних операторів***. У найпростішому випадку включення чи виключення специфічного терміна із запиту досягається додаванням префікса + чи - перед терміном. Наприклад, запит ***drugs +antiinflammatory -aspirin*** запустить пошук протизапальних препаратів, що відмінні від аспірину.

Більш гнучкі методи пошуку використовують інтуїтивно зрозумілі логічні оператори **AND**, **OR**, **NOT** (деякі пошукові системи дають змогу використовувати й інші оператори). Ці оператори в комбінації з дужками дозволяють задавати достатньо складні умови пошуку. Правила організації запиту можуть відрізнятись у різних пошукових системах, тому в кожному конкретному випадку варто ознайомитись із ними перед тим, як почати застосовувати складні варіанти пошуку.

Пошукові системи часто надають можливість обмежити пошук певними додатковими параметрами. Наприклад, пошук тільки в заголовках вебсторінок, пошук документів заданого формату (наприклад, лише PDF файлів) і навіть обмеження доменів, у яких проводиться пошук. Щоб уникнути нераціонального витрачання часу на пошук спеціалізованої інформації, варто починати його на добре відомих порталах, наприклад, присвячених хімікобіологічним наукам.

Спеціалізовані системи надають можливість крок за кроком знаходити потрібну спеціальну інформацію, проте відносно загальну інформацію, а, іноді, і спеціальні дані, можна знайти за допомогою загальних пошукових систем.

Безумовно, якість і повнота пошуку інформації залежить від повноти охоплення веб-ресурсів і технології пошуку, закладеної в системі. Дуже важливою є коректна побудова пошукового запиту користувачем, що вимагає певної практики.

Слід зазначити, що пошукові системи загального спрямування, наприклад, усім відомі системи Google, Yahoo,часто можуть бути корисними і для хіміків професіоналів. Тому з метою оптимізації витрат часу, пошук слід починати з пошукових систем загального спрямування, зокрема, Google.

При ***пошуку конкретної наукової публікації*** (при наявності вихідних даних) типи запитів можна розмістити у наступному порядку (за результативністю):

- DOI статті;
- повна назва статті мовою оригіналу;
- повна назва статті англійською мовою;
- повна або скорочена назва видання (журналу) (в цьому випадку система направить на сторінку даного видання, де треба буде знайти конкретний номер журналу за вихідними даними – зазвичай це можна зробити через архів номерів);

Якщо пошук не дав результату (особливо це стосується видань не англійською мовою), то на наступному етапі можна використовувати вже спеціалізовані пошукові системи. Найбільш повними і зручними у використанні являються:

Electronic Journals Library (Elektronische Zeitschriftenbibliothek EZB) (http://ezb.uni-regensburg.de/?lang=en) – електронна бібліотека журналів при бібліотеці університету Регенсбурга, надає посилання на майже 60 тис. безкоштовних журналів з різних галузей науки, техніки;

UCSF Library (University of California San Francisco) (https://www.library.ucsf.edu/journals/) – пошукова сторінка наукової бібліотеки університету Сан-Франциско, яка містить в своїй базі посилання на видання з різних напрямів, зокрема, 2556 онлайн-журналів з хімії;

Якщо пошук не дав результату і на цьому етапі, то вже треба використовувати спеціалізовані національні пошукові системи або на спеціалізовані ресурси відповідного напрямку.

При ***пошуку наукових публікацій за певним напрямом*** досліджень або за певними ключовими словами пошук також можна розпочати з пошукових систем загального спрямування, зокрема, Google, або Google Scholar, так як вони надають результати пошуку серед широкого кола джерел. На наступному етапі вже можна переходити до спеціалізованих тематичних баз даних і пошукових систем, докладно розглянутих у підрозділах 2.1.1, 2.1.2, 2.2.1–2.2.4.

Пошук ***наукометричних даних певного науковця*** краще одразу починати зі спеціалізованих систем ідентифікації науковців (див. підрозділ 2.2.4). Якщо відомий хоча б один із ідентифікаторів (ORCID номер, ResearcherID, Scopus Author ID), то пошук треба вести на відповідному ресурсі. Слід зазначити, що сторінка науковця у системі ORCID може містити всі його ідентифікатори, але це залежить від того, чи встановив сам науковець зв’язки між системами ORCID, Pablons і Scopus при реєстрації в системі ORCID.

Якщо ідентифікатори науковця невідомі, то пошук даних можна розпочати з системи ORCID, але слід враховувати, що прізвище науковця в цій сис-

темі може бути наведено будь-якою мовою, так як це залежить від самого науковця, який самостійно при реєстрації обирає мову написання прізвища і імені.

Пошук даних за прізвищем науковця можна також вести і в системах Scopus і Pablons, але в цих системах слід враховувати різні можливі варіанти написання прізвища та імені англійською мовою.

Якщо пошук не дав результату, то на наступному етапі можна шукати необхідні дані через пошукові системи загального спрямування, зокрема, Google або Google Scholar. Можна також спробувати знайти дані про певного науковця за його місцем роботи. Зараз зазвичай всі наукові і академічні установи розміщують на сторінках сайту своєї організації наукометричні дані або особисті сторінки співробітників-науковців.

ЛІТЕРАТУРА

1. ACD/ChemSketch. Version 2012 for Microsoft Windows. Drawing Chemical Structures and Graphical Images Tutorial. / Advanced Chemistry Development, Inc. 2013. – 156 p.
2. Bienz S. Short Manual to the Chemical Drawing Program ChemDraw / S. Bienz. University of Zurich. 2013. – 22 p.
3. MarvinSketch User's Guide. Available at: https://docs.chemaxon.com/display/docs/MarvinSketch+User%27s+Guide
4. Ракша О. В. Інформаційні технології у фізичній хімії: навчально-методичний посібник / О. В. Ракша. – Донецьк: ДонНУ, 2013. – 98 с.
5. Конспект лекцій (опорний) з дисципліни “Компьютерна хімія” / уклад.: М. Л. Кулігін. – Херсон: ХНТУ, 2013 – 80 с.
6. Винник О.Ф. Застосування програмного засобу ACD/ChemSketch (Freeware) 12.0 для написання хімічних формул та моделювання хімічних процесів. Навчальний посібник. / О.Ф. Винник, О.М. Свєчнікова, Т.Я. Грановська. – Харків, 2018. – 92с.
7. Фокін А. Г., Васильєва Л. В., Первухін В. М. Методичні вказівки до лабораторних робіт з дисципліни "Інформатика". "Основи роботи в мережі Internet".- Краматорськ : ДДМА, 2007. - 48 с.
8. Гужва В. М. Інформаційні системи і технології на підприємствах : навч. посібник.- К: КНЕУ, 2001. - 400 с.
9. Інформатика. Комп'ютерна техніка. Комп'ютерні технології : підручник / за ред. О. Пушкаря. - К: Академія, 2003. - 704 с.

Навчальне видання

КОМП'ЮТЕРНІ ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ В ХІМІЇ

Стислий конспект лекцій

для студентів спеціальності 102 «Хімія»
денної форми навчання

Укладач КОНОВАЛОВА Світлана Олексіївна

За авторським редагуванням

/2020. Формат 60 × 84/16. Ум. друк. арк.
Обл.-вид. арк. Тираж 5 пр. Зам. №

Видавець і виготівник
Донбаська державна машинобудівна академія
84313, м. Краматорськ, вул. Академічна, 72.
Свідоцтво суб'єкта видавничої справи
ДК №1633 від 24.12.2003