

Databáze v chemické a forenzní analýze

Tereza Uhlíková

April 13, 2026

Tereza Uhlíková

Ústav analytické chemie

skupina teoretické spektroskopie

místnost B4335

<https://web.vscht.cz/~uhlikovt/>

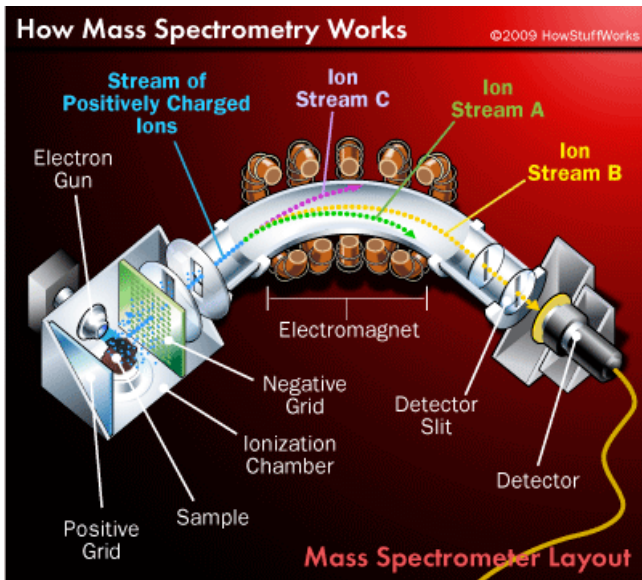
tereza.uhlikova@vscht.cz

O čem to bude?

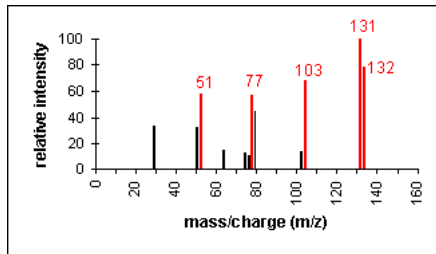
		Téma přednášky	Téma cvičení
5	17.3.	Základní principy AI	otázky - odpovědi - učení
6	24.3.	Použití ML	Algoritmy ML
7	31.3.	Chemické databáze	Analýza spekter pomocí databází
8	7.4.	Chemické databáze	Konkrétní použití ML - MS databáze
9	14.4.	Konkrétní použití ML, Forezní databáze - zkušková témata	
10	21.4.	dodělavky - opakování, ERA, normalizace, krystalografie	
11	28.4.	Ivan Raich	
12	5.5.	Ivan Raich	

Hmotnostní spektra

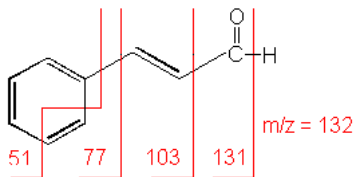
Princip hmotnostních spekter



Hmotnostních spektrum



Hmotnost jednotlivých fragmentů



3-Phenyl-2-propenal

C_9H_8O

MW = 132.16

- používá se ve spojení s plynovou chromatografií (GC) - nejprve se vzorek rozseparuje na GC, pak každá složka z GC jde do MS
- při zachování stejných ionizačních podmínek je MS velice dobře reprodukovatelná - hmotnost elementů je neměnná
- s knihovnou je potřeba porovnat celé MS spektrum, nejen konečnou molekulovou hmotnost
- v datech jsou zaznamenány pouze $[x;y]$ souřadnice - program spočítá pravděpodobnost dvou spekter
- člověk pak musí rozhodnout o správnosti

NIST knihovna <https://chemdata.nist.gov/>

<http://www.nist.gov/srd/nist1a.cfm>

Možno využít také tyto stránky - popis

<http://www.sisweb.com/software/nist-msms.htm>

demo verze <https://www.sisweb.com/software/ms/nist.htm#demo>

Online hledání sloučeniny, je-li v databázi

<http://www.sisweb.com/software/ms/nistsearch.htm>

Databáze MS peptidů

<http://chemdata.nist.gov/dokuwiki/doku.php?id=peptidew:mspepsearch>

Možnosti vyhledávání

- Nominální hmotnosti
- Přesné hmotnosti
- Vzorce/ nebo části
- Nějakáho píku
- Okrajové podmínky
- CAS číslo
- ID číslo
- NIST číslo
- Jméno/ i triviální

Wiley Registry

Wiley Registry of mass spectral data – už je placená
775500 hmotnostních spekter, 741000 chemických struktur za 6335 liber
<https://sciencesolutions.wiley.com/solutions/technique/gc-ms/wiley-registry-of-mass-spectral-data/>

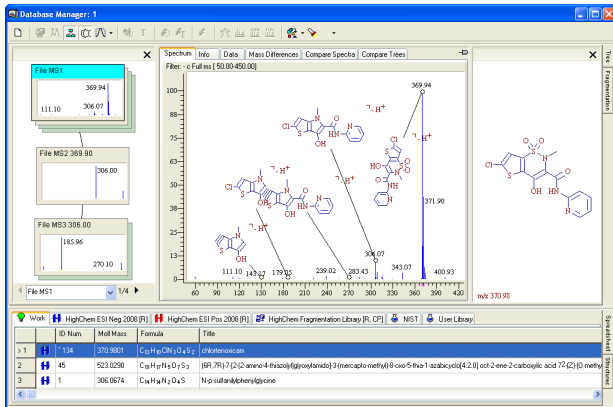
The screenshot displays the Wiley Registry software interface. The top menu bar includes File, Edit, Search, Predict, Show, Float, Clear, Settings, and Help. The interface is divided into several sections:

- Query area (red box):** Shows an empty mass spectrum plot with the x-axis labeled 'ppm' and values 140, 130, 120. The y-axis is labeled 'Arbitrary Unit' with values 0, 25, 50, 100. A chemical structure of 3-Chloro-aniline is shown to the right.
- Result area (blue box):** Shows a mass spectrum plot with several peaks. The x-axis is labeled 'ppm' and values 140, 130, 120. The y-axis is labeled 'Arbitrary Unit' with values 0, 25, 75. A chemical structure of 3-Chloro-aniline is shown to the right.
- Hitlist (green box):** A table with the following data:

Rank	Quality	Folder ID	Property
1/1	1000	1:FO0000000039	3-Chloro-aniline

Labels 'Query area', 'Result area', and 'Hitlist' are placed in colored boxes with arrows pointing to their respective sections. 'Icon bar' labels point to the toolbar icons in both the query and result areas.

HighChem Mass Frontier™; mzCloud <https://www.mzcloud.org/> Bratislava



Japonská massBank

<http://www.massbank.jp/?lang=en>

manuál

http://www.massbank.jp/manuals/MassBankUserManual_en.pdf

MassBank of North America

<http://mona.fiehnlab.ucdavis.edu>

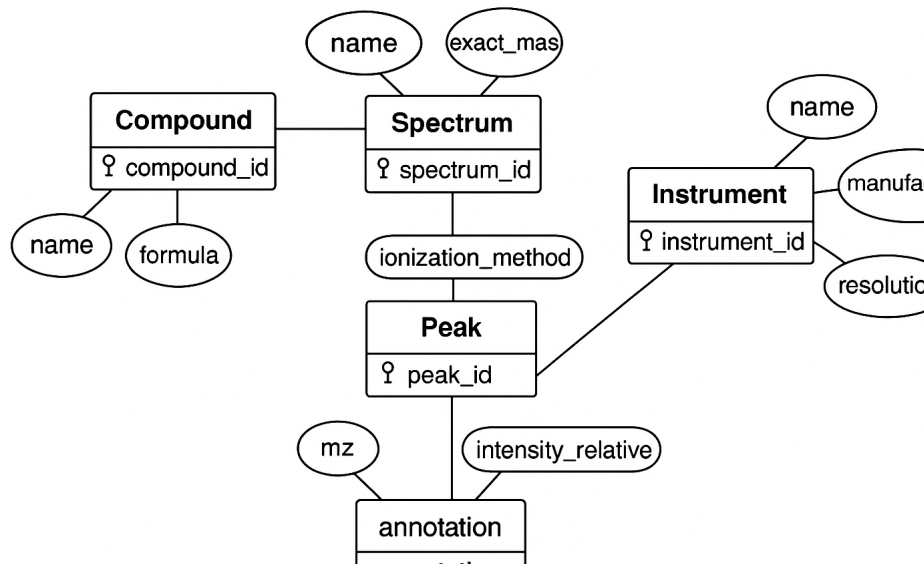
<http://mona.fiehnlab.ucdavis.edu/spectra/search>

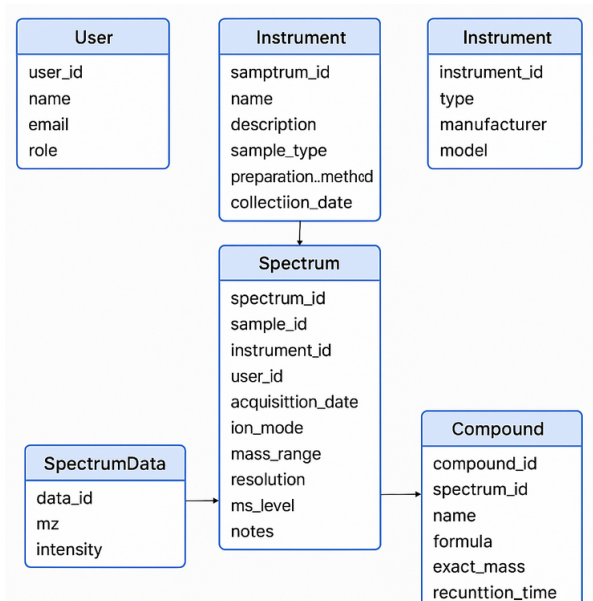
Vytvořme ERA pro databázi MS spekter v malé laboratoři

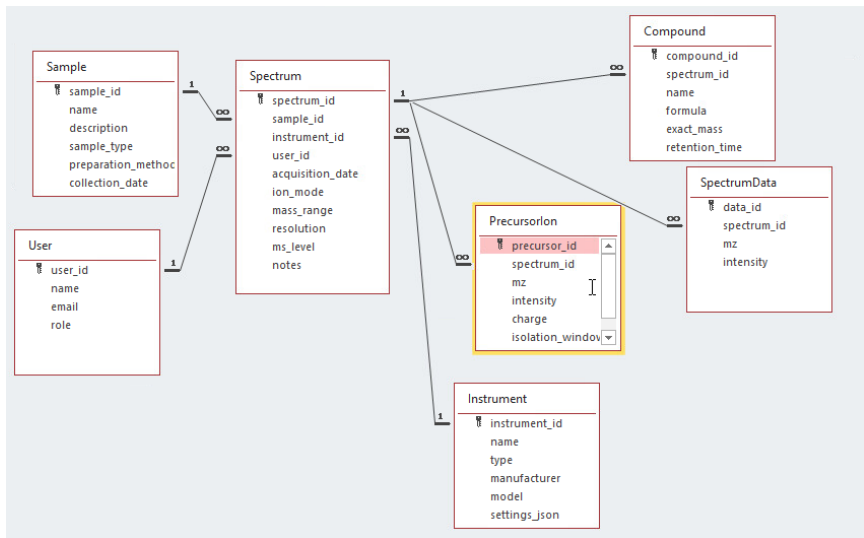
Entita	Příklady atributů
Molekula	název, sumární vzorec, CAS, přesná hmotnost
Spektrum	typ ionizace, přístroj, datum měření
Pík	m/z, relativní intenzita, anotace
Přístroj	typ, výrobce, rozlišení
Pracovník	jméno, pracoviště, kontakt, co dělal

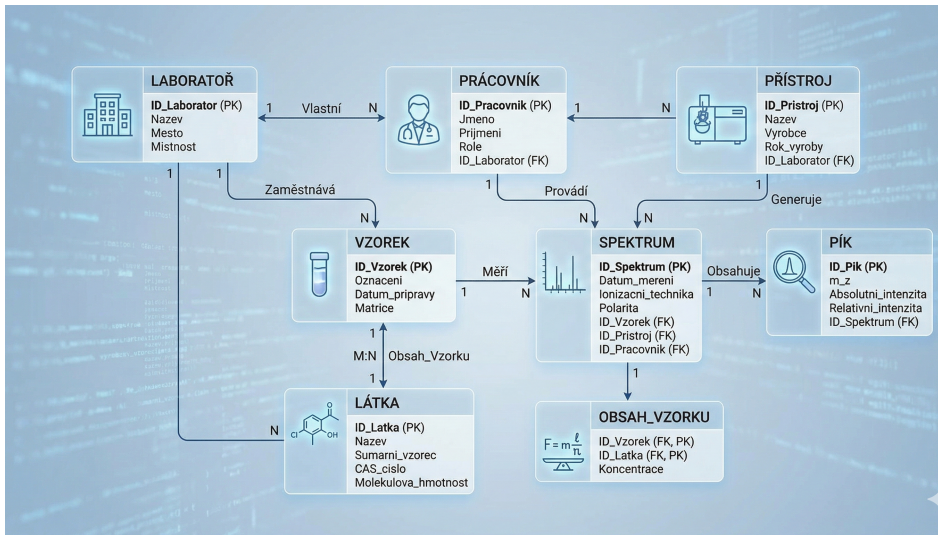
Relace

- MOLEKULA 1 : N SPEKTRUM = Jedna látka může mít více spekter (např. EI, ESI, různé fragmentace).
- SPEKTRUM 1 : N PÍK = Každé spektrum má mnoho píků.
- PŘÍSTROJ 1 : N SPEKTRUM = Jeden přístroj může měřit více spekter. Spektrum bylo naměřeno jedním přístrojem.









I textové modely vám vygenerují použitelný SQL kod.
Ten lze přímo vložit do např. MySQL/phpmyadnim prostředí.

- Deklarativní programovací jazyk
- Dotazovací jazyk - umožňuje ovládat databázi prostřednictvím příkazů – dotazů, za pomoci vyhledávacích operátorů
- Programovací jazyk, který se používá na operace s daty v relačních databázových SŘBD
- Příkazem SQL říkáme databázovému stroji, co chceme dělat (např. načíst data za určitých podmínek), ale neříkáme mu, jak má tento příkaz provést

čtyři základní skupiny:

- definice dat (DDL) - CREATE, ALTER, DROP, ... vytvářet, upravovat a mazat objekty databáze
- manipulace s daty (DML) - SELECT, INSERT, UPDATE, DELETE, ... získávat, ukládat a mazat data v databázi
- správa uživatelských rolí a práv (DCL) - GRANT, REVOKE, ... přidělení, odebrání oprávnění
- řízení transakcí (GTL) - START TRANSACTION, COMMIT, ROLLBACK ... hromadné provedení několika příkazů

Zásady pojmenovávání

- začínat písmenem
- dlouhé 1 až 30 znaků
- obsahovat jen A - Z, a - z, 0 - 9, _ (podtržítko), \$, a # (Nejsou! Case Sensitivní)
- nesmí být kopií jména dalšího objektu vlastněného stejným uživatelem
- název entity by měl odpovídat obsahu

Nejpoužívanější příkazy

CREATE DATABASE - vytváří novou databázi

ALTER DATABASE - upravuje databázi

CREATE TABLE - vytváří novou tabulku

ALTER TABLE - upravuje tabulku

UPDATE - upravuje data v databázi/tabulce

DELETE - maže data z databáze/tabulce

INSERT INTO - vkládá nová data do databáze/tabulky

DROP TABLE - maže tabulku

SELECT - výběr dat z databáze

```
CREATE TABLE Compound (  
  compound_id INTEGER PRIMARY KEY,  
  name TEXT,  
  formula TEXT,  
  exact_mass REAL,  
  cas_number TEXT );
```

```
CREATE TABLE Instrument (  
  instrument_id INTEGER PRIMARY KEY,  
  name TEXT,  
  manufacturer TEXT,  
  resolution INTEGER );
```

```
INSERT INTO Compound VALUES
```

```
(1, 'Benzene', 'C6H6', 78.04695, '71-43-2'), (2, 'Toluene', 'C7H8',  
92.06260, '108-88-3');
```

```
INSERT INTO Instrument VALUES
```

```
(1, 'Q-Exactive', 'Thermo Fisher', 140000), (2, 'GC-MS 5977B', 'Agilent',  
5000);
```

Najdi všechna spektra látky „Benzene“

```
SELECT S.spectrum_id, S.ionization_method, S.date_measured
FROM Spectrum S
JOIN Compound C ON S.compound_id = C.compound_id
WHERE C.name = 'Benzene';
```

Najdi všechny píky příslušející k určitému spektru

```
SELECT mz, intensity_relative
FROM Peak
WHERE spectrum_id = 1
ORDER BY mz;
```

Vyhledávání podle spektra

Krok 1: Celé spektrum v **SpectrumData**

Krok 2: Přidat **Index** na **mz**

```
CREATE INDEX idx_mz_search ON SpectrumData (mz);
```

rychlejší vyhledávání

```
SELECT spectrum_id FROM SpectrumData WHERE mz BETWEEN  
445.33 AND 445.35;
```

lepší je použít složený index

```
CREATE INDEX idx_mz_spectrum ON SpectrumData (mz,  
spectrum_id);
```

uchovávání celého naměřeného spektra (BLOB, TEXT)

```
ALTER TABLE Spectrum ADD COLUMN spectrum_blob BLOB;
```

-- nebo

```
ALTER TABLE Spectrum ADD COLUMN spectrum_json TEXT;
```

```
-- Users INSERT INTO User (name, email, role) VALUES  
( 'Alice Nguyen', 'alice@lab.org', 'Technician'),  
( 'Bob Silva', 'bob@lab.org', 'Researcher'),  
( 'Dr. K. Jordan', 'jordan@lab.org', 'PI');
```

Do Accessu není možné :-)

Musí se vkládat řádek po řádku:

```
INSERT INTO Person VALUES (1, "Amir");  
INSERT INTO Person VALUES (2, "Sofia");  
INSERT INTO Person VALUES (3, "Aya");
```

Vkládání SQL kodu do Accessu

Máme SQL skript (CREATE TABLE + INSERT)

Postup v Microsoft Access:

- Otevři Access → vytvoř novou databázi
- Jdi na: Vytvořit → Návrh dotazu
- Zavři okno „Přidat tabulku“
- Přepni do: Zobrazení SQL
- Vlož svůj SQL kód

Problém:

- Access má jinou SQL syntaxy než např. MySQL / PostgreSQL
- SQL z MySQL nejde spustit přímo
- datové typy nesedí
- cizí klíče a constraints dělají problémy

AUTO_INCREMENT → AUTOINCREMENT; INT → INTEGER
VARCHAR → TEXT; odstranit ENGINE=InnoDB apod.

1. Import přes mezikrok

nahraj data do: MySQL nebo SQLite

exportuj do: CSV

2. Import do Accessu

V Accessu: Externí data → Nový zdroj dat → Z textového souboru
vyber CSV

projdi průvodce

access romana sovakova - youtube videa

např. <https://www.youtube.com/watch?v=XAwV193fwdA>

<https://edu.gcfglobal.org/en/access2013>

Karel kohout access - youtube videa

např. <https://www.youtube.com/watch?v=LaNf1iEeznY>

<https://www.youtube.com/watch?v=mAMBOXzltng> validation email address