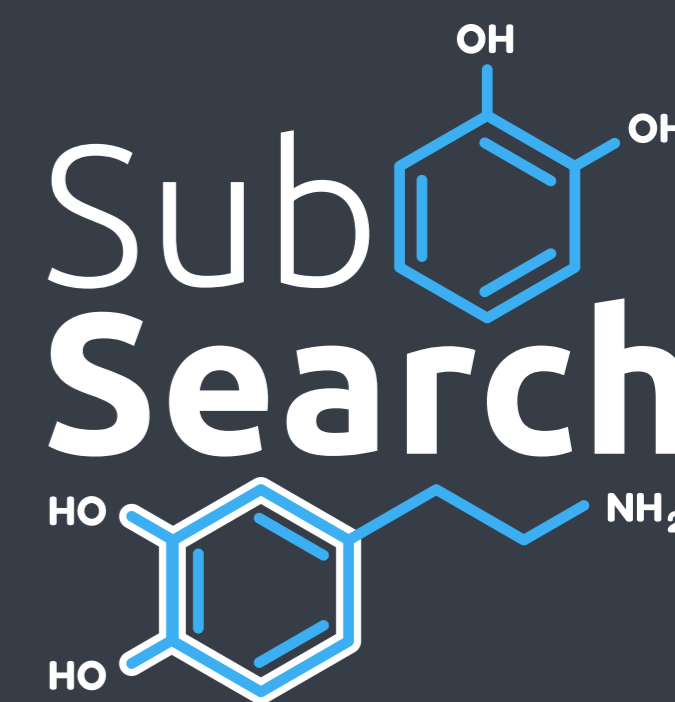


# Podštruktúrne vyhľadávanie v databázach chemických látok

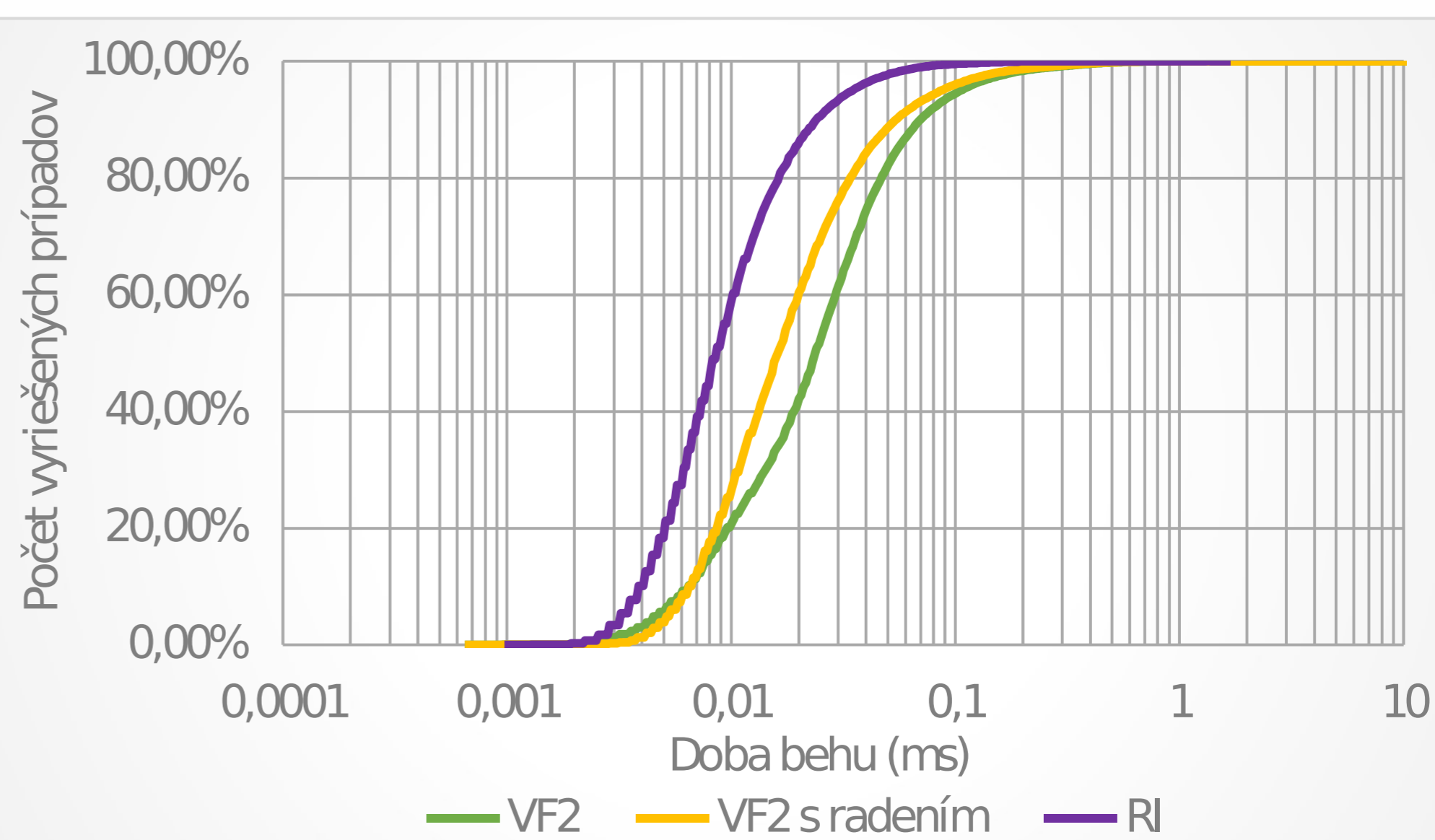
Ivan Ševčík, xsevci50@stud.fit.vutbr.cz



## Zhrnutie

Práca sa zaoberá vytvorením systému pre podštruktúrne vyhľadávanie v databázach chemických látok obsahujúcich desiatky miliónov záznamov. Zo svojej podstaty je podštruktúrne vyhľadávanie časovo náročná operácia. Navrhnuté riešenie využíva kombináciu molekulárnych odtlačkov a rýchleho algoritmu pre hľadanie izomorfného podgrafu. Rovnako dôležité bolo navrhnúť spôsob uloženia štruktúrnych dát v systéme, kde bolo zvolené uloženie serializovanej reprezentácie do databáze MongoDB.

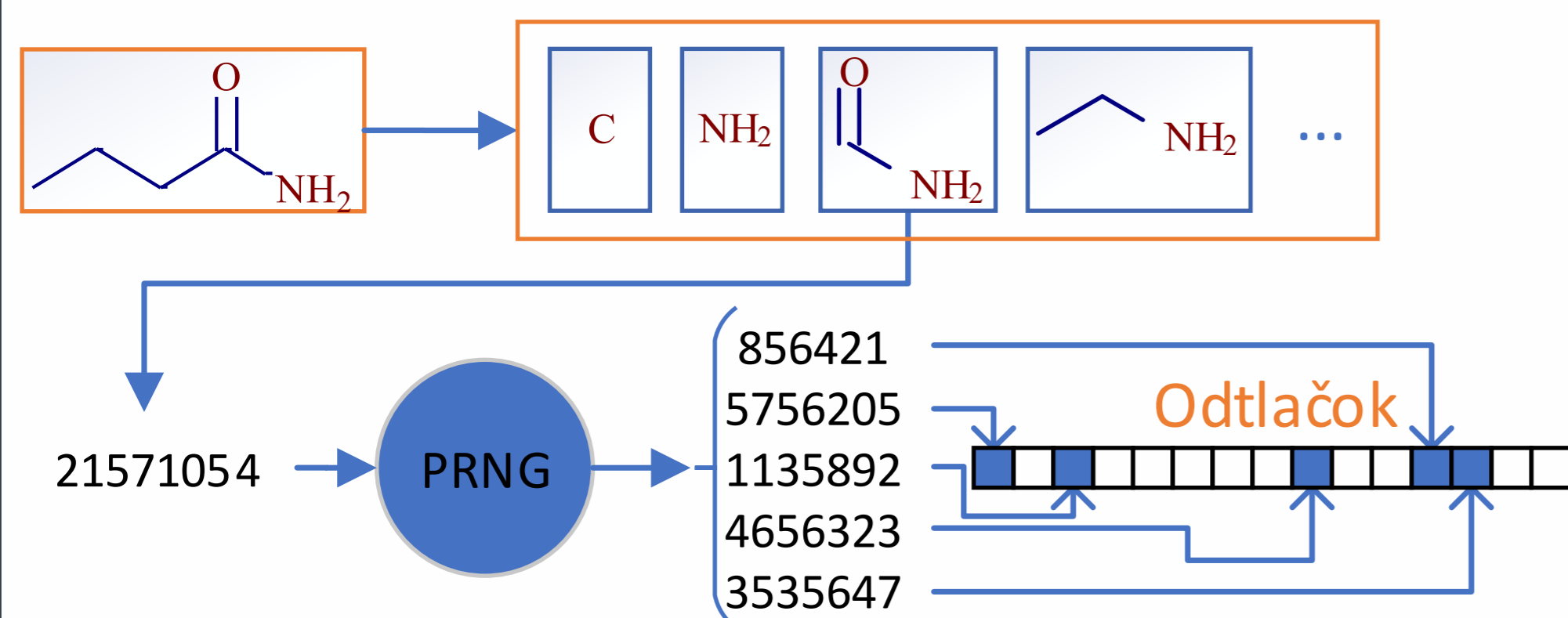
## Porovnanie algoritmov



Algoritmus	Doba behu ( $\mu$ s)	Najrýchlejší (#)
VF2	37.014	734941
VF2 s radením	29.088	96473
RI	12.817	5641108
Best	12.259	

## Princíp odtlačku

Podštruktúrny odtlačok z knižnice RDKit. Presnosť odtlačku je približne 60%, čo môže viesť k zredukovaniu počtu prehľadávaných štruktúr o viac ako 99%.



## Porovnanie uloženia

Uloženie štruktúrnych dát do stromovej štruktúry v súborovom systéme a NoSQL databáze MongoDB umožňujúcej získať viacero záznamov súčasne.

Úložisko	Doba získania dát ( $\mu$ s)	
NTFS	211.936	
MongoDB	1	234.376
	10	35.108
	100	13.408

## Architektúra systému

