Computer-Assisted Markush Structures Curation from Patent Documents

David Deng, Árpád Figyelmesi
Motivation

Knowing the chemical space covered by competitors’ patents is essential for successful drug discovery.

This information can be useful for:
- Idea generation
- Lead candidates selection
- Drug design
- Patent claims construction
Challenges

- The existing databases are optimized for patent searching
- Automatic processing quality is not good enough
- Manual processing and analysing is time consuming and needs special expertise
- No tools for creating, editing and visualizing complex Markush structures
ChemAxon’s Technology “Pieces”

- *Document to Structure* recognizes English, Chinese and Japanese documents dramatically speeds up the extraction process
- Markush Representation, Search and Enumeration
- Structure Checker and Markush Validation guarantees the high quality of extracted information
- Markush Editor helps to draw complex Markush structures
Name to Structure

- Support for many nomenclatures (common, drug names, ...)
- IUPAC names used for exemplified structures and R-group fragments
- Essential to extract chemical information from patents

- English (2008, Marvin 5.1)
- Chinese (2013, Marvin 5.12)
- Japanese (2014, Marvin 6.3)
实施例4

\[ \text{III-4N-对硝基苄基-4-苯甲酰基-4-哌啶醇盐酸盐} \]

先按通法一的合成及后处理方法制备4-苯甲酰基甲基-4-哌啶醇(II) 1.7mol和4-苯甲酰基甲基-4-哌啶醇(III) 1.90(8.8mmol)和无水K$_2$CO$_3$3.53g(25.6mmol)置于无水丙酮(60ml)中回流反应3小时，然后减压蒸馏，得到白色晶体2.68g，收率81.9%。

元素分析：C$_{20}$H$_{22}$N$_2$O$_4$ HClH$_2$O(理论值 %：C 55.15, H 6.28，N 6.95，Cl 8.95) 实验值 %：C 58.52，H 6.28，N 7.04，Cl 8.95)

$^1$HNMR(DMSO-d$_6$): 1.80-2.10(m, 4H, 哌啶H), 2.20-2.35(m, 4H, CH$_2$), 2.50-2.80(m, 4H, CH$_2$), 3.13-3.80(m, 6H, 哌啶H), 4.25(s, 2H, PhCH$_2$), 7.25(s, 1H, 芳香H)。
Markush Technology

ChemAxon
Structure Checker

- Drawing errors
- OSR Scanning errors
- Inconsistent representations
Markush Editor

- Editing complex patent Markush structures
- Hierarchical representation of fragments’ relationships
- Visualization of nesting view, preview
- Editing separately the individual fragments
- Integrated structure checker
- Available as a desktop application and as an integratable component
Markush Editor Demo
Markush Composer

- Creating Markush structures from set of compounds
- Available as an experimental feature in Markush Editor
ChemCurator

- Compound and Markush editor component
- Annotated documents
- Drag and drop structures from the document
- Connection between the document and extracted data
- Markush validation against the examples
- Support for multi-display environment
- Will be available as a desktop application in Q4
Workflow

- IP experts can represent the chemical space
- Chemical representation is comprehensible for Medicinal Chemists
- High quality project specific database
- New opportunities, less risk, faster communication
General Document Curation

Usable for extracting specific structures
- Scientific journal articles
- Internal company reports
- Chemical patents

Wizard to automatically detect relevant structures
- Exclude fragments, molecular weight, substructure...
Future plans

Product launch on Markush Session at ChemAxon US UGM, September 25-26, Cambridge, MA

In the pipeline:
- Overlap analysis
- Non-hit visualization
- Making Composer smarter
- Improving Markush representation
Acknowledgment

Daniel Bonniot
Árpád Figyelmesi
Gábor Botka
David Deng
Péter Kovács