HOW TO IDENTIFY NEW CHEMISTRY IN ASIAN PATENTS

East meets West 2018
Challenges

• Language barriers
  Machine translation of chemical names is extremely hard with traditional NLP techniques

• Volume of the data
  • Volume of newly published patents
  • Volume of prior art documents
Name to Structure conversion

5-{2-ethoxy-5-[(4-methylpiperazin-1-yl)sulfonyl]phenyl}-1-methyl-3-propyl-1H,6H,7H-pyrazolo[4,3-d]pyrimidin-7-one

シルデナフィル

Sildenafil

5-{2-エトキシ-5-[(4-メチルピペラジン-1-イル)スルホニル]フェニル}-1-メチル-3-プロピル-1H,6H,7H-ピラゾロ[4,3-d]ピリミジン-7-オン

 Viagra

139755-83-2

伟哥
Identification of chemical entities

Identification of chemical entities

Identification of chemical entities

Identification of chemical entities

Identification of chemical entities

Identification of chemical entities

Identification of chemical entities

Identification of chemical entities

Identification of chemical entities

Identification of chemical entities

Identification of chemical entities

Identification of chemical entities

Identification of chemical entities

Identification of chemical entities

Identification of chemical entities

Identification of chemical entities

Identification of chemical entities

Identification of chemical entities

Identification of chemical entities

Identification of chemical entities

Identification of chemical entities

Identification of chemical entities

Identification of chemical entities

Identification of chemical entities

Identification of chemical entities

Identification of chemical entities

Identification of chemical entities

Identification of chemical entities

Identification of chemical entities

Identification of chemical entities

Identification of chemical entities
Published patents
Case study

Find new potentially patentable drug analogs and new ring systems
• Filter drugs from ChEMBL
• Search analogs in GDB-13
• Analyze overlap with SureChEMBL

## Case study

|          | GDB-13 analogue | Tversky | SureChEMBL Tversky | GDB-13 analogue | Tversky | SureChEMBL Tversky | GDB-13 analogue | Tversky | SureChEMBL Tversky | GDB-13 analogue | Tversky | SureChEMBL Tversky | GDB-13 analogue | Tversky | SureChEMBL Tversky | GDB-13 analogue | Tversky | SureChEMBL Tversky | GDB-13 analogue | Tversky | SureChEMBL Tversky |
|----------|-----------------|---------|-------------------|-----------------|---------|-------------------|-----------------|---------|-------------------|-----------------|---------|-------------------|-----------------|---------|-------------------|-----------------|---------|-------------------|-----------------|---------|-------------------|-----------------|---------|-------------------|
|          |                 | 0.23    | 0.23              |                 | 0.23    | 0.23              |                 | 0.23    | 0.23              |                 | 0.23    | 0.23              |                 | 0.23    | 0.23              |                 | 0.23    | 0.23              |                 | 0.23    | 0.23              |                 | 0.23    | 0.23              |
|          |                 | 0.06    | 0.06              |                 | 0.09    | 0.06              |                 | 0.08    | 0.08              |                 | 0.05    | 0.06              |                 | 0.06    | 0.07              |                 | 0.05    | 0.05              |                 | 0.05    | 0.01              |                 | 0.05    | 0.05              |
|          |                 | 0.23    | 0.24              |                 | 0.26    | 0.26              |                 | 0.26    | 0.26              |                 | 0.26    | 0.26              |                 | 0.26    | 0.26              |                 | 0.26    | 0.26              |                 | 0.26    | 0.27              |                 | 0.26    | 0.27              |
|          |                 | 0.09    | 0.13              |                 | 0.11    | 0.13              |                 | 0.08    | 0.01              |                 | 0.01    | 0.01              |                 | 0.09    | 0.10              |                 | 0.07    | 0.07              |                 | 0.07    | 0.07              |                 | 0.07    | 0.07              |
|          |                 | 0.23    | 0.24              |                 | 0.24    | 0.24              |                 | 0.24    | 0.24              |                 | 0.24    | 0.24              |                 | 0.24    | 0.24              |                 | 0.24    | 0.24              |                 | 0.24    | 0.24              |                 | 0.24    | 0.24              |
|          |                 | 0.01    | 0.09              |                 | 0.09    | 0.07              |                 | 0.07    | 0.07              |                 | 0.01    | 0.09              |                 | 0.08    | 0.08              |                 | 0.08    | 0.07              |                 | 0.08    | 0.07              |                 | 0.08    | 0.07              |
|          |                 | 0.31    | 0.31              |                 | 0.31    | 0.32              |                 | 0.32    | 0.32              |                 | 0.33    | 0.33              |                 | 0.33    | 0.33              |                 | 0.33    | 0.33              |                 | 0.33    | 0.33              |                 | 0.33    | 0.33              |
|          |                 | 0.09    | 0.05              |                 | 0.09    | 0.10              |                 | 0.05    | 0.12              |                 | 0.09    | 0.08              |                 | 0.09    | 0.09              |                 | 0.09    | 0.09              |                 | 0.09    | 0.09              |                 | 0.09    | 0.09              |
Proposed workflow

• Extract chemical structures automatically from newly published Asian patent documents

• Compare with existing databases e.g. SureChEMBL

• Identify automatically novel ring systems and new compound series by using large scale clustering techniques and overlap analysis
THANK YOU

afigyelmesi@chemaxon.com