Assisting Data Retrieval With a Drug Knowledge Graph

Romain LELONG\(^{a,b,1}\), Badisse DAHAMNA\(^{a,b}\), Romain LEGUILLON\(^{a}\), Julien GROSJEAN\(^{a,b}\), Catherine LETORD\(^{a,b}\), Stéfan J.DARMONI\(^{a,b}\), Lina F. SOUALMIA\(^{b,c}\)

\(^{a}\)CHU Rouen, Department of Biomedical Informatics, F-76000 Rouen, France
\(^{b}\)LIMICS U1142, Sorbonne Université, Paris, France
\(^{c}\)Normandie Univ, UNIROUEN, TIBS-LITIS EA 4108, F-76000 Rouen, France

12 octobre 2021

1. romain.lelong@gmail.com
Prescription orders integration

Rouen University Hospital (RUH)

RUH Health Data Warehouse EDSaN

1,452,616 prescription orders
6,978,586 atomic prescriptions
2011 → August 2021

Information retrieval?
Prescriptions metadata & Information Retrieval

Patient information
id, age, gender, birthdate

Stay information
id, entry date, leaving date, units

prescription information
id, date

List of prescribed drugs
drug label, details, UCD codes

Lucene indexes ⇒ allow the search by metadata
Prescriptions metadata & Information Retrieval

**Patient information**
- id, age, gender, birthdate

**Stay information**
- id, entry date, leaving date, units

**Prescription information**
- id, date

**List of prescribed drugs**
- drug label, details, UCD codes

---

**Search for prescription orders based on what drug were prescribed?**

UCD = Common Dispensing Unit

⇒ Encoding and billing of administered drugs in France

⇒ Good quality of data

**Selecting drugs**

= 

**Selecting UCD codes**
Drug Knowledge Graph

Medical Subject Heading descriptors
Anatomical Therapeutic Chemical (ATC) concepts
International Non proprietary Names (INNs)
Virtual drugs from the MédicaBase
Pharmacological roots
Pharmacological specialties
Drug components
Drug components groups
UCD codes

131,277 vertices
703,807 edges

HeTOP²

Health Terminology/Ontology Portal
> 90 terminologies or ontologies
> 2 million concepts

2. https://hetop.eu
Exploitation of the knowledge graph

Matching the input term against a vertex label

Traversing the graph and selecting paths leading to UCD codes

weighting & classification of paths

Paths distributed over 7 relevance classes (C1 to C7)

Paths weighting

Vertex 1
\[ w_{r1} \]
Vertex 2
\[ w_{r2} \]
Empirically assigned with a hospital pharmacist
Vertex 3
\[ w_{rn} \]
Vertex n

\[ w_{path} = w_{r1} + w_{r2} + \cdots + w_{rn} \]
Initial vertex matching the input term

Input term

Slider to adjust tolerance

Retrieved UCD codes
88 terms randomly drawn among the possible types of vertex.

System

Only the paths/UCD codes from the first three non-empty relevance classes were considered. More than 100,000 paths were retrieved.

Hospital pharmacist evaluation

Attribution of a score to each path/UCD code from 1 to 3: 1 unsatisfactory 2 could be improved 3 consistent result.
### Result & Discussion

<table>
<thead>
<tr>
<th>Type of term</th>
<th>$C_1$</th>
<th>$C_2$</th>
<th>$C_3$</th>
<th>$C_4$</th>
<th>$C_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drug composition</td>
<td>2.71</td>
<td>1.08</td>
<td>1.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drug composition groups</td>
<td>2.71</td>
<td>3.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>MeSH Descriptors</td>
<td>2.94</td>
<td>1.49</td>
<td>1.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>INN</td>
<td>2.94</td>
<td>1.49</td>
<td>1.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>virtual drugs</td>
<td>3.00</td>
<td>1.23</td>
<td>1.28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medical indications</td>
<td>3.00</td>
<td>3.00</td>
<td>2.82</td>
<td>1.52</td>
<td></td>
</tr>
<tr>
<td>ATC Code</td>
<td>2.25</td>
<td>3.00</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pharmacological specialties</td>
<td>2.50</td>
<td>1.59</td>
<td>1.89</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pharmacological roots</td>
<td>3.00</td>
<td>1.84</td>
<td>1.05</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1. Overall, average scores of paths tends to decrease with level of relevance but exceptions can be found;
2. Highest score of 3 (or close to 3) reached for several types of terms;
3. Variability of the best possible relevance class.
Does the drug knowledge graph assist the user in selecting drugs of interest?

**Yes**: The ranking of resulting UCDs codes were overall congruent with the expert judgment.

**But**: Some inconsistencies remain.

⇒ A refinement of the weights assigned to the edges of the graph is necessary!
Thank you for your attention

Questions?