

## Background

- Necessity of Standard Term Conversion
  - Use of different terminologies in each institute and country
- Standard Vocabulary in OMOP: RxNorm, SNOMED, LOINC, etc.
  - Drug vocabulary in OMOP: ATC, RxNorm, etc.

## Purpose

- Development of the candidate concepts recommendation algorithm
  - Extraction of RxNorm descendant (candidates) using ATC code relation
  - Ranking of each candidate according to a similarity algorithm

## Method

- Standard concept recommendation algorithm
  - Provision of lists of most similar concepts (candidates) of RxNorm in order using Jaccard index
    - Jaccard index: Measurement of a similarity between a source name of the source code and a normalized name (concept name) of RxNorm
  - Concept candidates are ranked according to each similarity

- Process of the recommendation algorithm
  - Generation of a source name and an ATC code of the code (Input)
  - Algorithm Execution
    - Extraction of candidates of the concepts through OMOP vocabulary relation using the ATC code of the source code
    - Measurement of each string similarity between the source name and each concept name of the candidates based on the Jaccard index
  - Ranking of each candidate according to similarity results (Output)

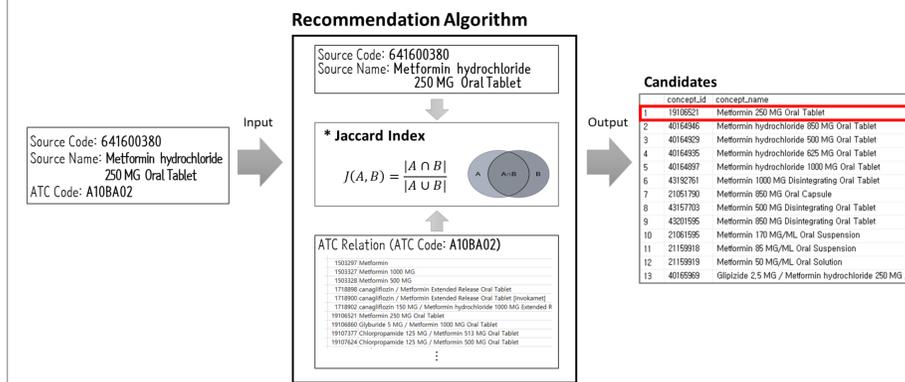


Figure 1. Example (source code: 644904330) of recommendation algorithm

- Extraction of RxNorm descendant from a source code (644904330)
  - Concept ID of ATC code 'A10BA02'

Table: CONCEPT		
concept_code	concept_id	...
A10BA02	21600747	...

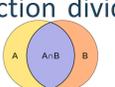
- The ATC code maps to Concept ID of RxNorm

Tables: CONCEPT_RELATIONSHIP, CONCEPT					
concept_id_1	concept_id_2	concept_name	relationship_id	concept_class_id	vocabulary_id
21600747	1503297	Metformin	Maps to	Ingredient	RxNorm

- Descendant of Concept ID '1503297'

Tables: CONCEPT_ANCESTOR, CONCEPT					
concept_id	concept_name	ancestor_concept_id	domain_id	vocabulary_id	concept_class_id
1503297	Metformin	1503297	drug	RxNorm	Ingredient
1503327	Metformin 100 MG	1503297	drug	RxNorm	Clinical Drug Comp
1503328	Metformin 500 MG	1503297	drug	RxNorm	Clinical Drug Comp
1718898	canagliflozin / Metformin Extended Release Oral Tablet	1503297	drug	RxNorm	Clinical Drug Form
1718900	canagliflozin / Metformin Extended Release Oral Tablet [Invokamet]	1503297	drug	RxNorm	Clinical Drug Form

Candidate rows: 930

- Jaccard Index for measuring the source name and each candidate names
  - Measurement of similarity between sample sets (e.g., A, B)
  - Calculation of the size of their intersection divided by the size of their union as follows:  $J(A, B) = \frac{|A \cap B|}{|A \cup B|}$ 

  - The result of similarity: 0 – 1 (1: the highest similarity)
- e.g. Jaccard similarity measurement of sample sets
  - A (source name): Metformin hydrochloride 250 MG Oral Tablet
  - B (concept name): Metformin 250 MG Oral Tablet
  - $|A \cap B| = \{\text{Metformin, 250, MG, Oral, Tablet}\} = 5$
  - $|A \cup B| = \{\text{Metformin, hydrochloride, 250, MG, Oral, Tablet}\} = 6$
  - $J(A, B) = 5/6 = 0.83$

## Experimental Results

- There are various string similarity methods: Jaccard index, Levenshtein, etc.
- Levenshtein selection for comparing the performance of Jaccard index
  - String metric for measuring the edit distance between two sequences
  - Minimum number of operations required to change one word into the other
    - Operations: consisting of insertions, deletions or substitutions of a single character, or transposition of two adjacent characters
  - e.g. Levenshtein measurement of sample sets
    - A (source name): Metformin hydrochloride 250 MG Oral Tablet
    - B (concept name): Metformin 250 MG Oral Tablet
    - Distance Result: 14 (0 is same)
- Experimental Environments
  - Number of Sample source codes : 100 and 300 (random generation)
  - To measure accuracy of Levenshtein and Jaccard Index
    - Use of golden standard data completed through medical experts

- Ratio of accurate mapping versus Top-k occurrence (Figure 2)

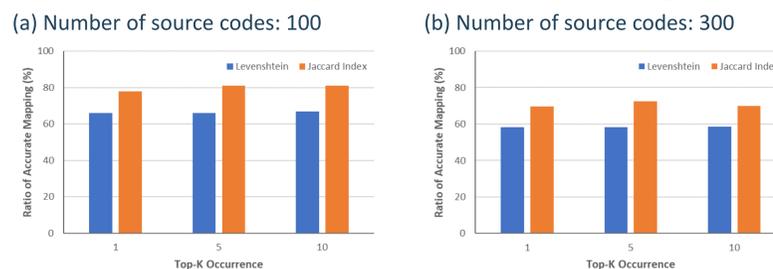


Figure 2. Ratio of accurate mapping versus Top-k occurrence (Blue color: Levenshtein, Orange color: Jaccard index)

- Jaccard Index's accuracy is higher than Levenshtein
  - In case (a), Jaccard index is more than 13% accurate as compared to Levenshtein; in case (b) Jaccard index is more than 12% accurate

## Conclusions

- Proposal of standard concept recommendation algorithm
  - Extraction of RxNorm descendant (candidates) from an ATC code of a source code
  - Ranking of each candidate based on Jaccard index
    - Jaccard index's accuracy is higher than Levenshtein as 100 and 300 source codes are generated
    - Jaccard index is more than average 12% accurate compared to Levenshtein
- Acknowledgment: This research was supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number : HI16C0992)