Knowledge Discovery Based on the Identification of Relationships Among Medical Concepts by Clustering Candidate Descriptors

Shizuka Yamazaki[a], Mariko Iwasawa[b], Shin-ichi Nakayama[b], Natsuo Onodera[b], Yusuke Tanigawara[a]

[a] *Department of Pharmacy, School of Medicine, Keio University, Japan*

[b] *Graduate School of Library, Information and Media Studies, University of Tsukuba, Japan*

Introduction

- The discovery of hidden connections within the medical literature could give new and plausible knowledge.

Objective

- This study proposes and examines a new methodology for hypothesis generation and knowledge discovery using cluster analysis.

Methods

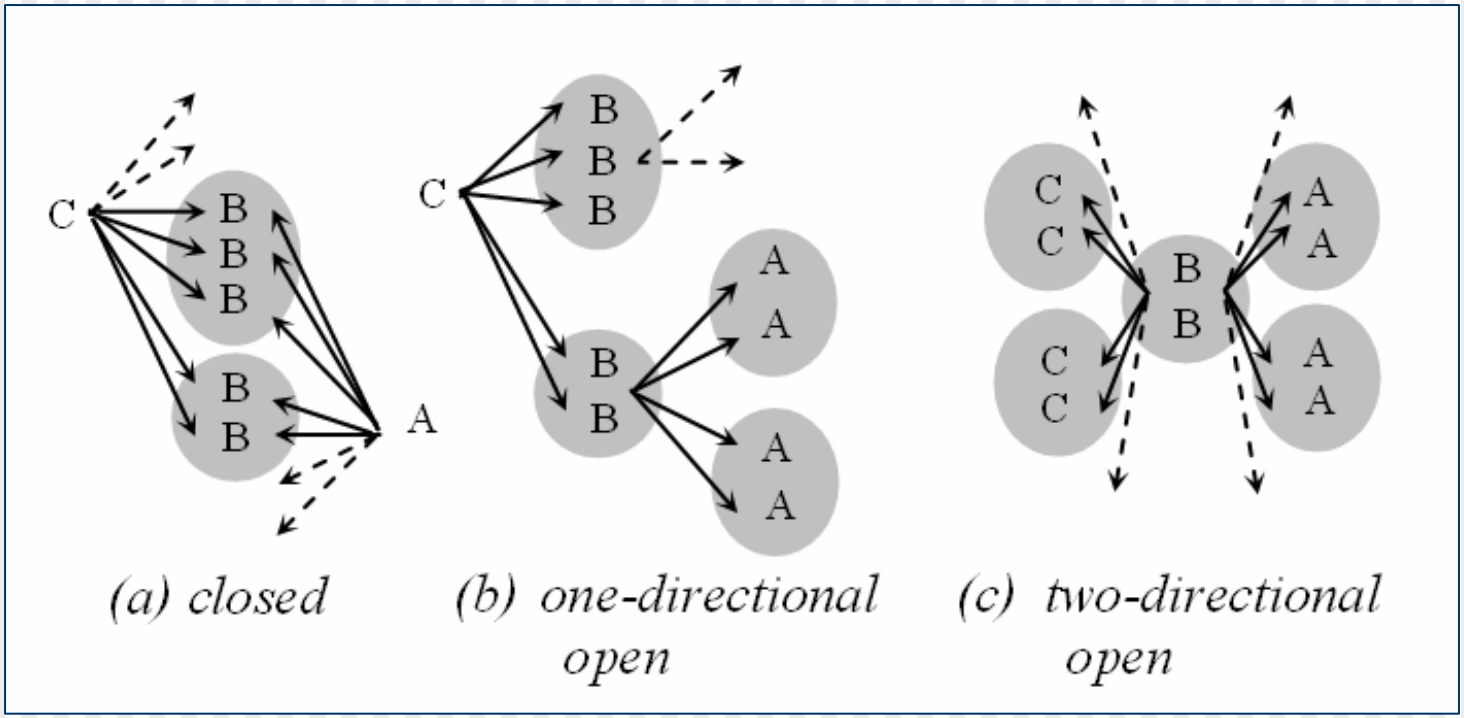
Background

- Don R. Swanson (1986), in his the ABC model, showed that unknown relationships between a chemical substance (A) and a disease (C) could be found by identifying a concept (B) that connects A and C in the literature .
- Closed and one-directional open discovery processes were reported by Marc Weeber et al. (2001).
- Most prior works used ranking based on term frequency. Such methodology may bring on only a part of the concepts.

We propose grouping candidate MeSH terms by means of hierarchical clustering for obtain knowledge overview.

Methods

Discovery process



(a) closed *(b) one-directional open* *(c) two-directional open*

We constructed the term-document matrix for clustering candidate MeSH terms.

Methods

(a) Closed discovery process

The B terms were MeSH terms that co-occurred in the document sets A and C.

i. Creating document sets A and C (B-terms collection)
ii. B-terms grouping

(b) One-directional open discovery process

The B terms were MeSH terms in the set C. These B terms were grouped and the B terms in each group were used to extract set A from MEDLINE.

i. Creating document set C (B-terms collection)
ii. B-terms grouping
iii. Creating document set A (A-terms collection)
iv. A-terms grouping

(c) Two-directional open discovery process

The A and C terms were MeSH terms in the set B

i. Creating document set B (A-terms, C-terms collection)
ii. A-terms, C-terms grouping

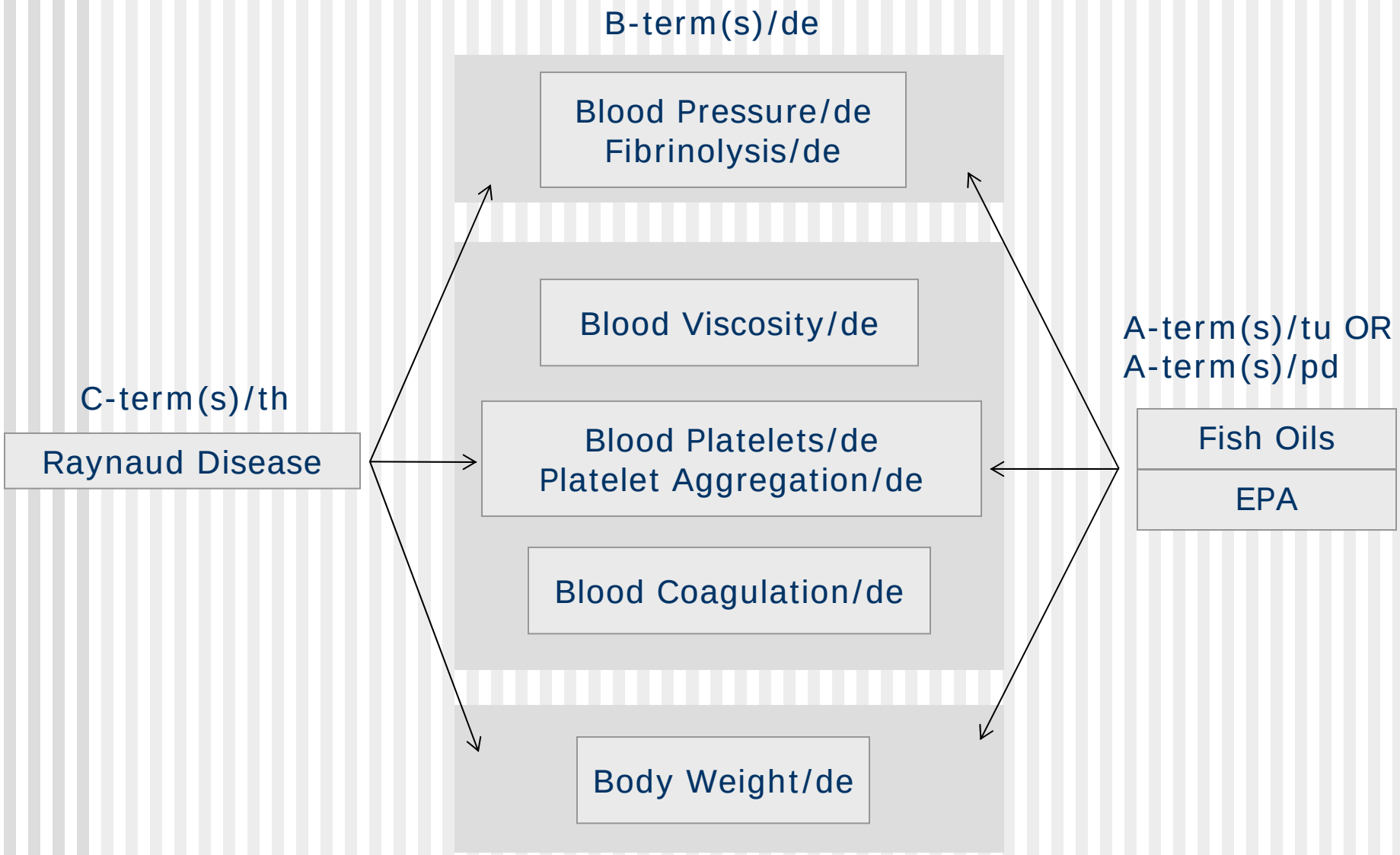
Methods

- We searched for particular terms with the relevant subheading:

```
C-term(s)/th [therapy]
   - B-term(s)/de [drug effect]
      - A-term(s)/tu [therapeutic use]   OR
        A-term(s)/pd [pharmacology].
```

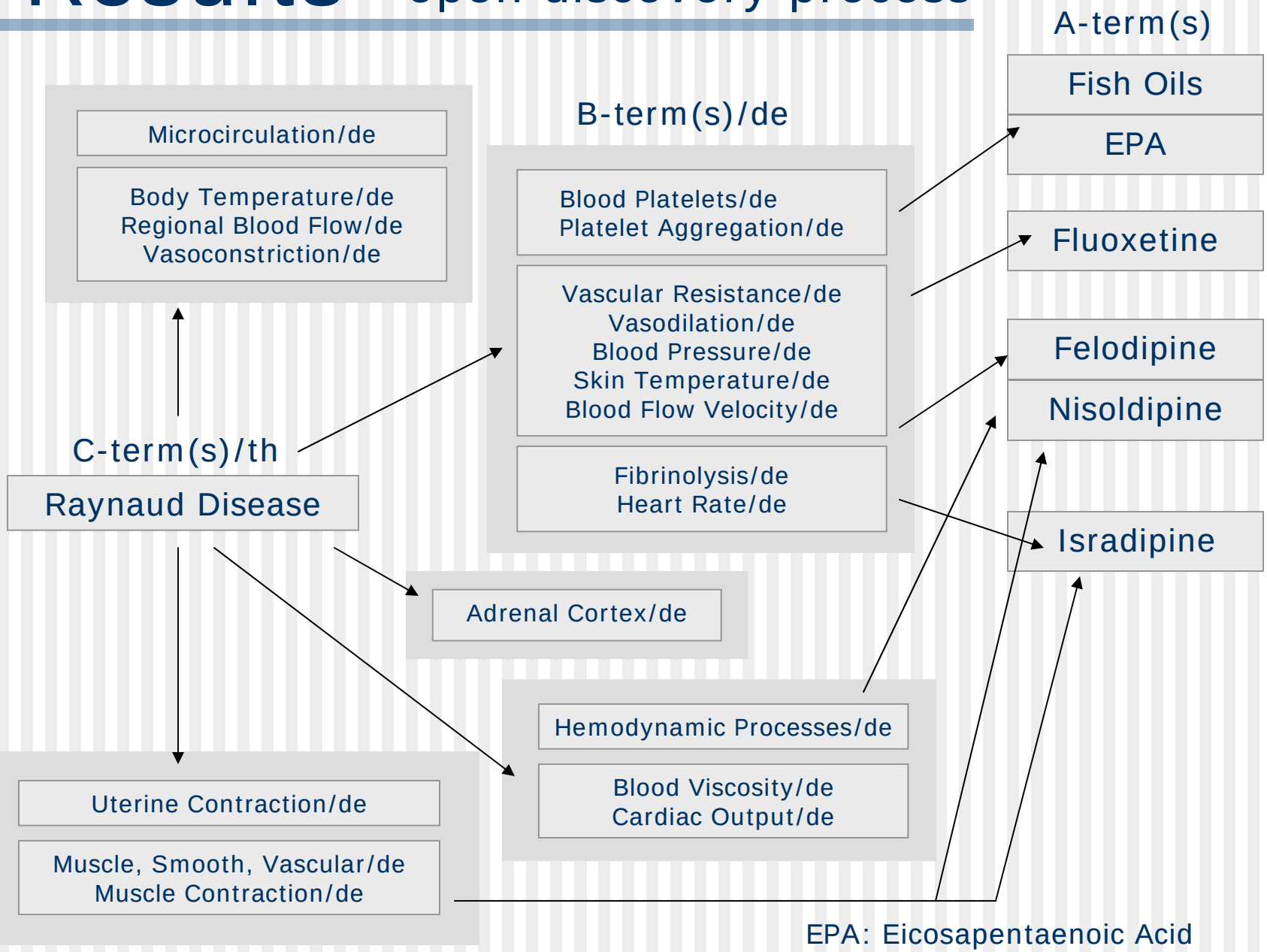
- In each process, the stop list of MeSH terms used was similar to that applied in Swanson's method (2006).
- Our discovery processes were applied in an attempt to identify relationships between fish oil and Raynaud's disease from MEDLINE documents as an experimental test of Swanson's hypothesis.
  Database: PubMed http://www.ncbi.nlm.nih.gov/PubMed/
  ("1983/01/01"[PDAT] : "1985/12/31"[PDAT])

Results Closed discovery process



Results

One-directional open discovery process



EPA: Eicosapentaenoic Acid

Results

One-directional open discovery process

B-term(s)/de

Blood Platelets/de
Platelet Aggregation/de

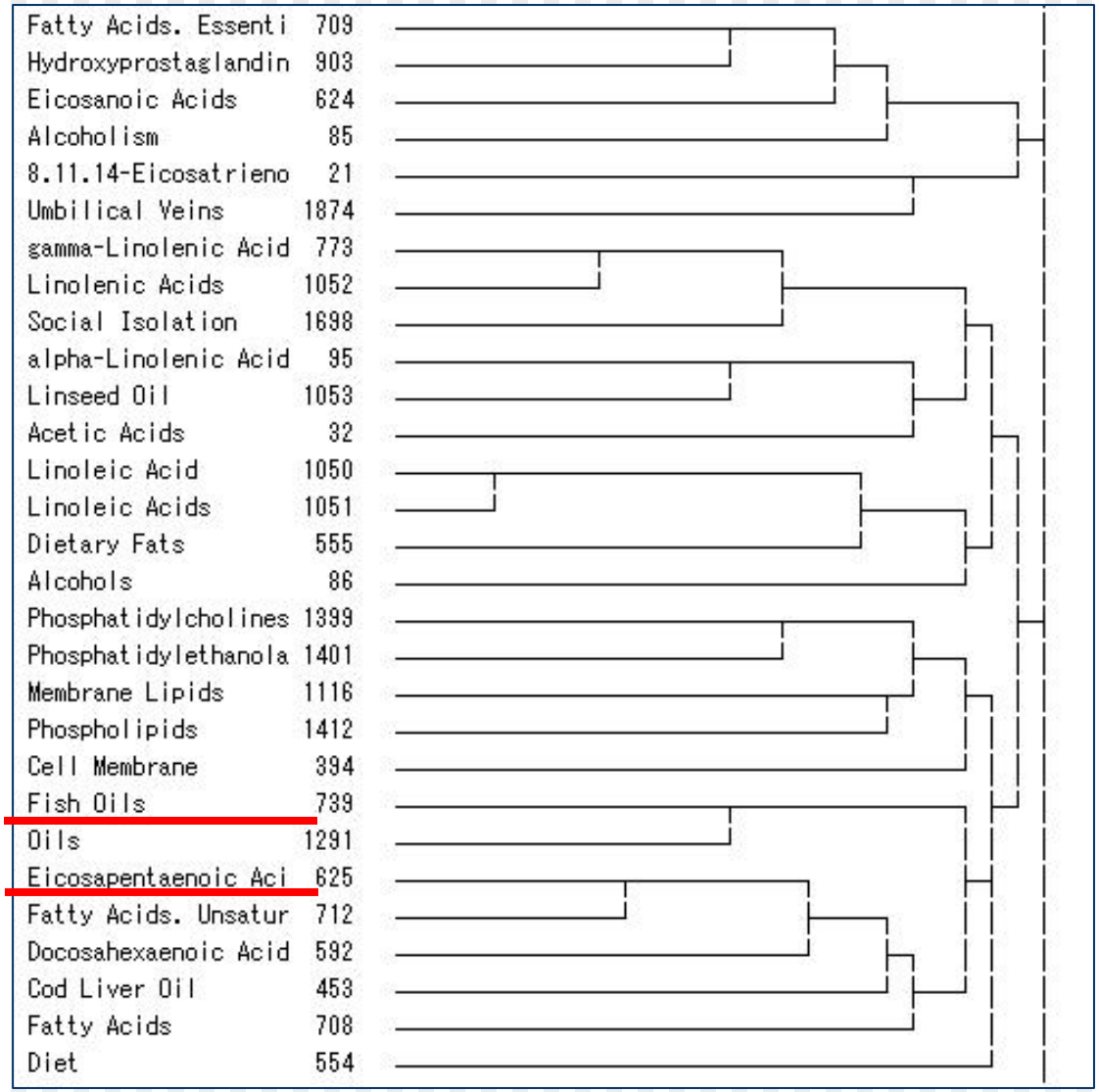
Vascular Resistance/de
Vasodilation/de
Blood Pressure/de
Skin Temperature/de
Blood Flow Velocity/de

Fibrinolysis/de
Heart Rate/de

A-term(s)

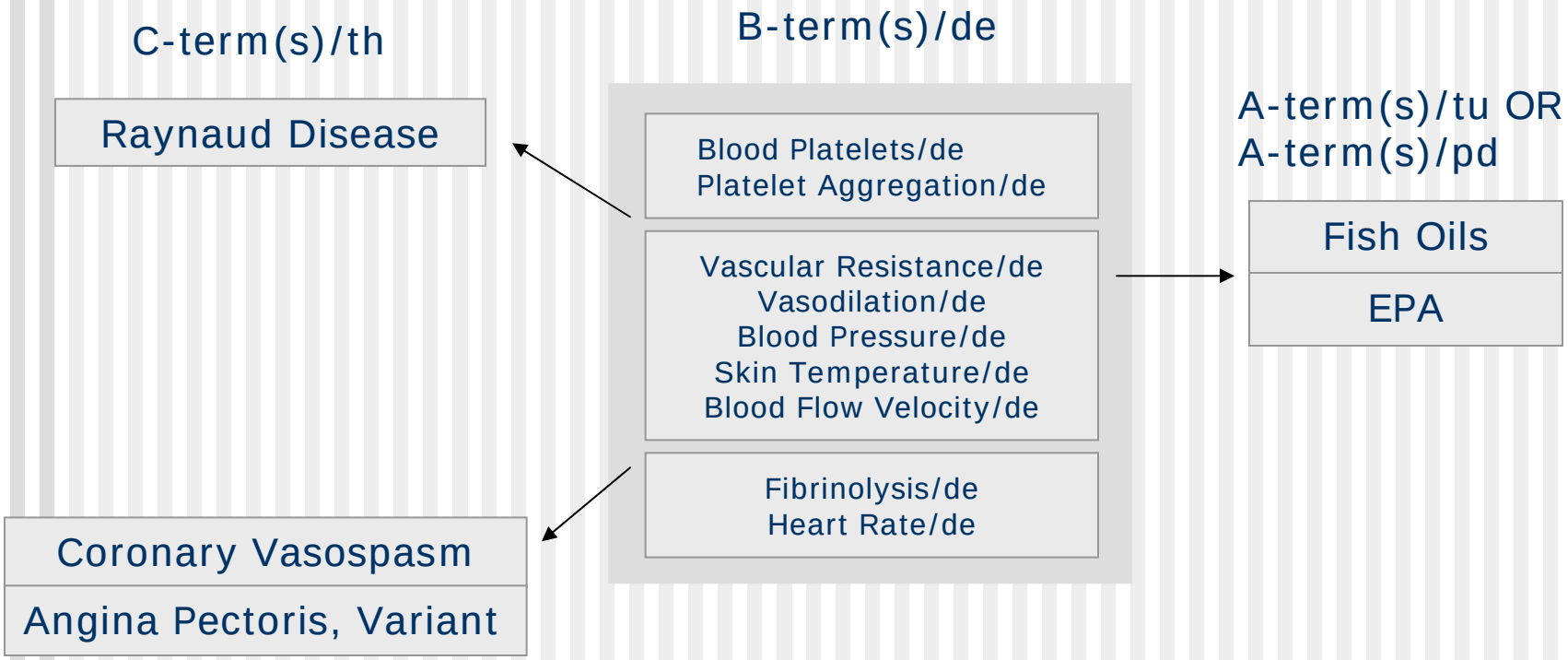
Fish Oils

EPA



Results

Two-directional open discovery process



Discussion and conclusion

- By grouping candidate descriptors using cluster analysis, we succeeded in detecting invisible connections between medical concepts.
- It is important to obtain knowledge on relationships between various phenomena observed in clinical practice. All relationships are not always direct connections, however.
  Our new methodology was effective in identifying hidden connections between medical concepts.
- The resulting knowledge overview could be helpful in clinical practice.

References

[1]Swanson DR. Fish oil, Raynaud's syndrome, and undiscov-ered public knowledge. Perspectives in Biology and Medi-cine. 1986. 30(1), 7-18.

[2]Weeber M, Klein H, de Jong-van den Berg LTW, and Vos R. Using concepts in literature-based discovery: Simulating Swanson's Raynaud-fish oil and migraine-magnesium dis-coveries. Journal of the American Society for Information Science and Technology. 2001: 52(7), 548-557.

[3] Swanson DR. Smalheiser NR. Torvik VI, Ranking indirect connections in literature-based discovery: The role of medical subject headings. Journal of the American Society for Information Science and Technology. 2006: 57(11), 1427-1439.

Address for correspondence

Shizuka Yamazaki
Department of Pharmacy, School of Medicine, Keio University, Japan
E-mail: yshizu@sc.itc.keio.ac.jp