

Thirteenth International Workshop on
Juris-informatics (JURISIN 2019)

Exploring Relevant Parts between Legal Documents using Substructure Matching

Takahiro Komamizu

Kazuya Fujioka

Yasuhiro Ogawa

Katsuhiko Toyama



Ordinances and Rules in Local Governments

(OR documents for short)

- Regulations for social lives
- In Japan, about a million in 1,788 local governments
- In wide range



harmful contents for kids



drunk-driving



environmental pollution

Drafting OR Documents

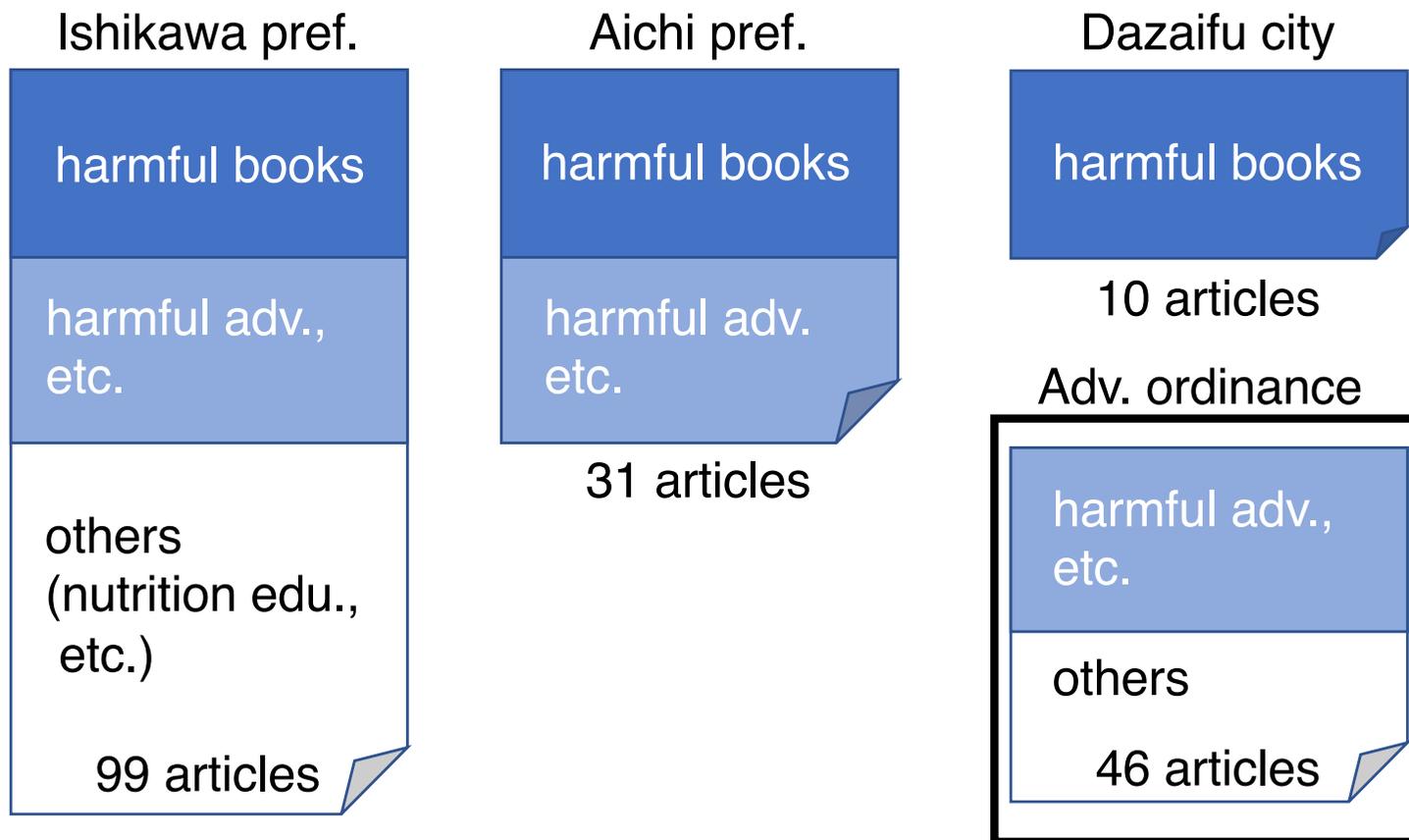
- Officers draft according to social situations.
 1. Search relevant existing OR documents,
 2. imitate them for the first draft, and
 3. modify the content for the situations.
- Questions
 - How officers search *relevant* OR documents?
 - Heuristics (e.g., those of “similar” governments, popular ones, etc.)
 - How they choose which *parts* of the documents for imitation?
 - On manual

What we can do?

- OR document search \approx document similarity search
 - Vectorization (BoW, TF-IDF, Doc2vec, etc.)
 - Topic modeling (LSI, LDA, etc.)
 - Similarity computation (cosine similarity, Tversky index, etc.)
- Parts determination for imitation
 - Idea: different parts (e.g., sentences) are missing in the drafting OR document.
 - *diff* can be a choice.
 - However, there can be too many different sentences.
 - Even one character difference.
 - Moreover, irrelevant provisions would be included.

Various Granularity of Provisions

- Dependent on individual local governments.
 - e.g., Protection of young persons ordinances



“Same” provision = Same content? – No

- Different governments have different contents.
 - Different structure and ordering, missing contents, etc.

Paragraphs about council in Landscape ordinances

Ami			Shichigashuku		
#A	#P	Article Title (Content desc.)	#A	#P	Article Title (Content desc.)
23	1	Establishment	11	1	Council (Establishment)
24	1	Deliberation matters	11	2	Council (Deliberation matters)
25	1	Counsel	11	3	Council (Counsel)
26	1	Organization (#Committee)	12	1	Organization (#Committee)
26	2	Organization (Conditions)	12	3	Organization (Conditions)
26	3	Organization (Temporary com.)			
27	1	Term of service (Basic)	12	2	Organization (Term of service)
27	2	Term of service (Substitute)			
27	3	Term of service (Temporary com.)			
28	1	Chairperson (Election)			
28	2	Chairperson (Chairperson)			
28	3	Chairperson (Vice chairperson)			
29	1	Convention (Summons)			
29	2	Convention (Resolution)			
30	1	Section			

Objective

- Intuition

- Missing contents in “same” provisions are helpful to determine which parts to imitate.

Ami			Shichigashuku		
#A	#P	Article Title (Content desc.)	#A	#P	Article Title (Content desc.)
23	1	Establishment	11	1	Council (Establishment)
24	1	Deliberation matters	11	2	Council (Deliberation matters)
25	1	Counsel	11	3	Council (Counsel)
26	1	Organization (#Committee)	12	1	Organization (#Committee)
26	2	Organization (Conditions)	12	3	Organization (Conditions)
26	3	Organization (Temporary com.)			

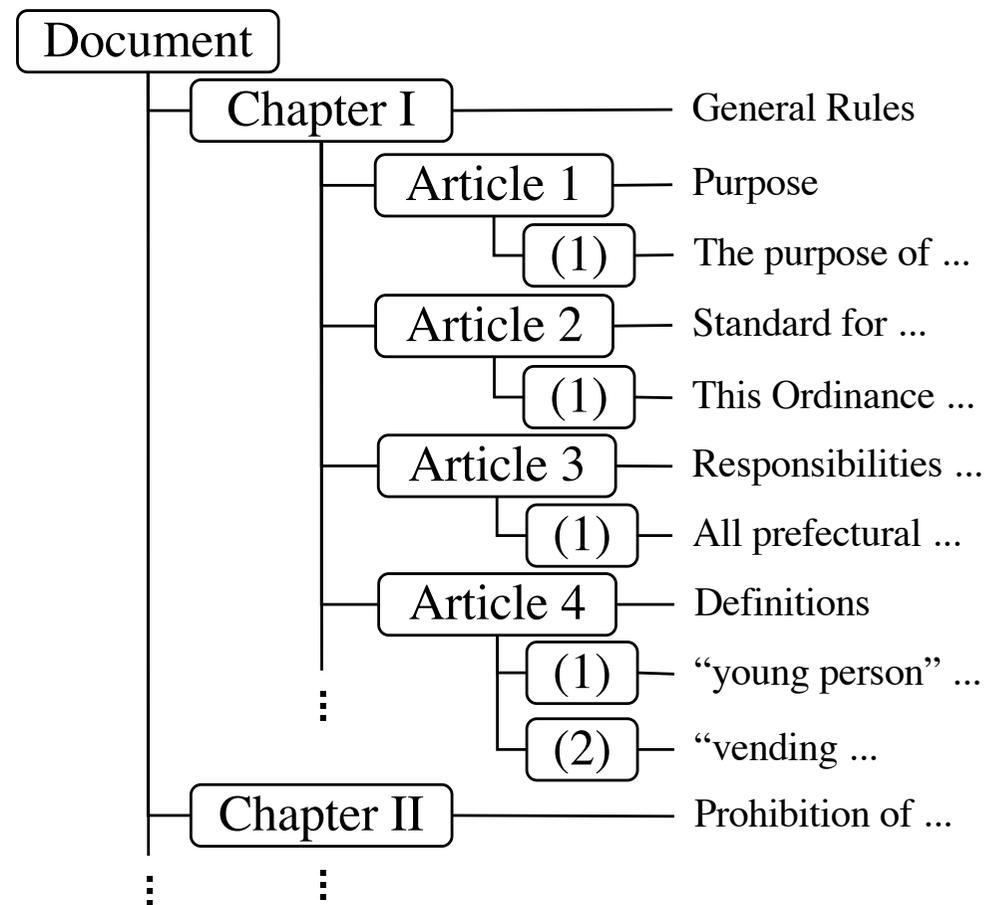
- Questions

- Which parts are corresponding with “same” provisions?
- How to determine missing contents?

Idea: Substructure matching by regarding OR documents as tree-structured data

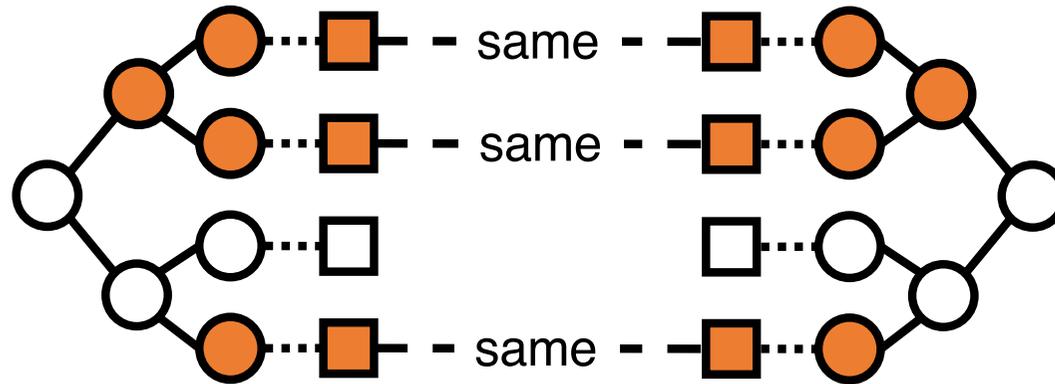
- OR documents as tree-structured data
 - Subtree is regarded as provision

Chapter I General Rules
Article 1 Purpose
 (1) The purpose of this Ordinance is to protect young persons by preventing activities which ...
Article 2 Standard for Operation
 (1) This Ordinance shall apply to the minimum extent necessary to achieve the purpose in ...
Article 3 Responsibilities of Protection
 (1) All prefectural residents are required to protect young persons from the environment that ...
Article 4 Definitions
 (1) “young person” means a person under 18 years of age.
 (2) “vending machine” means an equipment for the sale of goods, which is capable of selling the ...
...
Chapter II Prohibition of acts tends to impede the sound upbringing of young persons
...

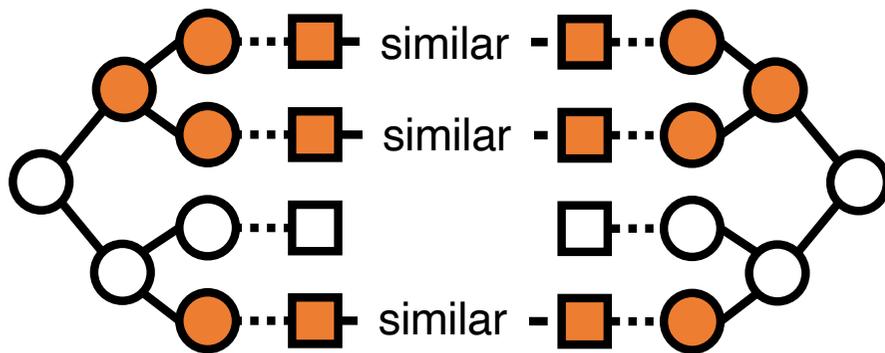


Idea: Substructure Matching

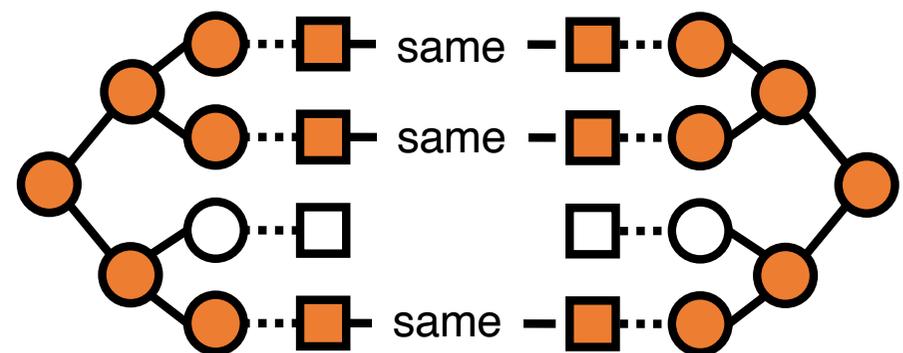
- Same provision: subtree pair having same contents



- Relaxations: similarity matching



Content matching relaxation



Structural matching relaxation

Proposed measurement: Matching Ratio

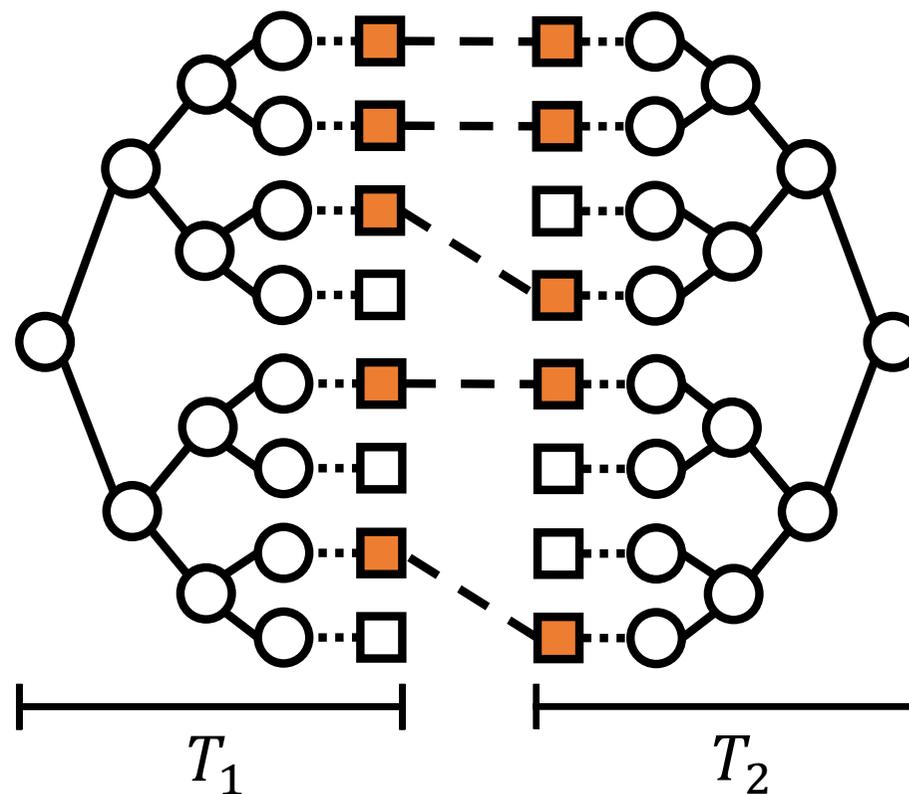
- Intuition:
the more matched contents,
the more similar provisions

$$R(T_i, T_j, M_i, M_j) = \sqrt{\frac{|M_i|}{|S_i|} \cdot \frac{|M_j|}{|S_j|}}$$

M_i : set of **matched contents** in T_i
 S_i : set of **leaf nodes** in T_i

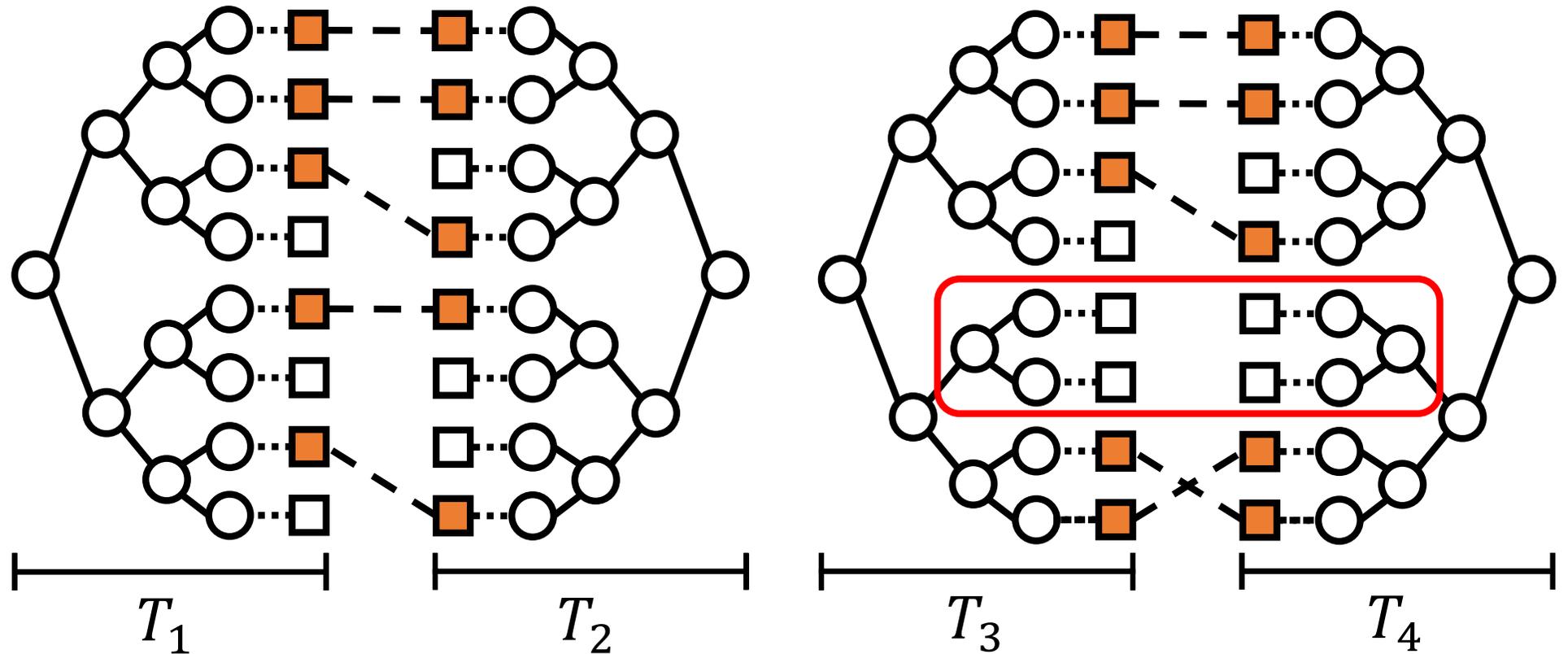
- Example on the right

- $R(T_1, T_2, M_1, M_2) = \sqrt{\frac{5}{8} \cdot \frac{5}{8}} = 0.63$



Same Matching Ratio, but ...

- Matching ratio cannot distinguish the following.

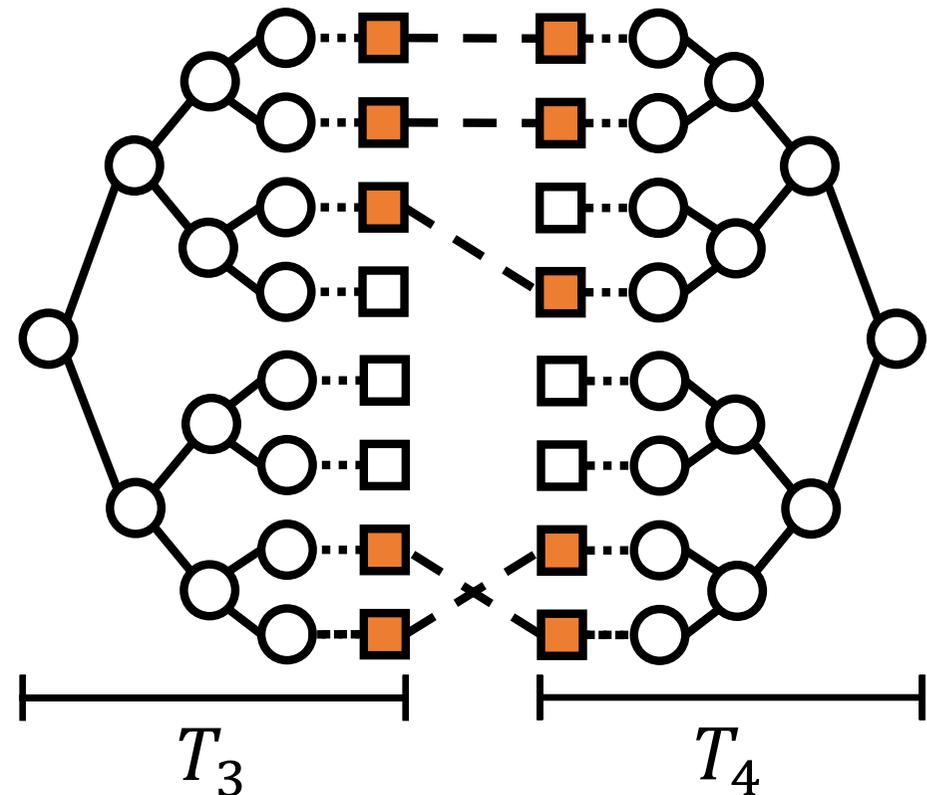


Proposed measurement: Provision commonality

- Intuition:
the larger number of consecutive non-matched leaf nodes, the more different contents are included.

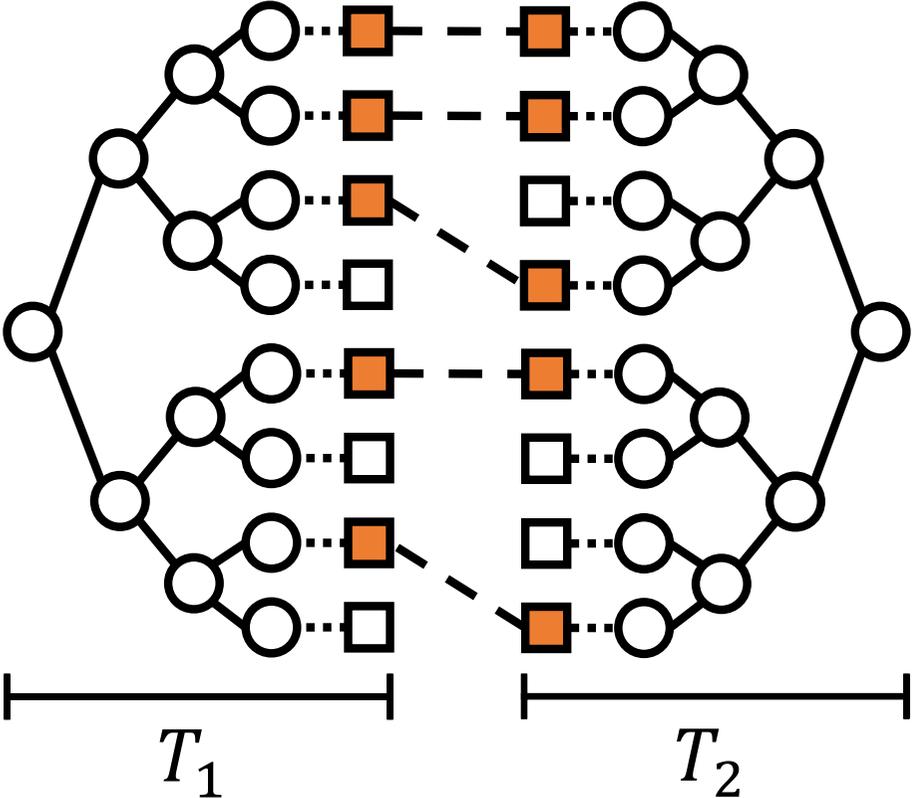
$$C(T_i, T_j, M_i, M_j) = \sqrt{\frac{1}{L(T_i, M_i)+1} \cdot \frac{1}{L(T_j, M_j)+1}}$$

$L(T_i, M_i)$: the longest consecutive non-matched leaf nodes in T_i

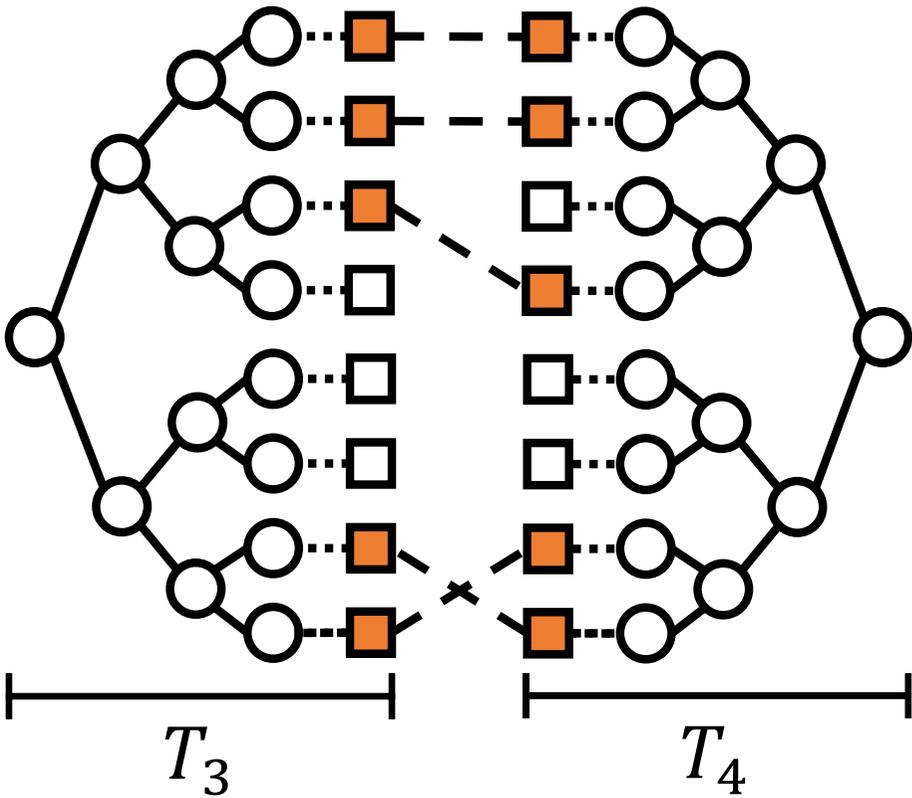


Provision commonalities

$$C = \sqrt{\frac{1}{1+1} \cdot \frac{1}{2+1}} = 0.41$$



$$C = \sqrt{\frac{1}{3+1} \cdot \frac{1}{2+1}} = 0.29$$



Search Algorithm (see paper for detail)

- Given:
 - Two OR documents: T_i, T_j
 - Contents matching: M
- Find:
 - Vertex pairs $P \subseteq T_i.V \times T_j.V$
s.t. linear combination of $R(\cdot)$ and $C(\cdot) > \text{threshold}$
 - P is maximal to avoid redundancy
 - Vertex pair $(v, u) \notin P$ s.t. ascending vertex pairs of v, u are in P
- Idea
 - Bottom up manner
 - Eliminate candidate pairs if they cannot be in P .

Evaluation

- Settings

- Survey by Ito [6]

- Inclusions of specific classes of provisions (e.g., council and prohibitions) in **articles** of the landscape ordinances.

- Examination

- Check whether discovered pairs belong same class.
 - Baseline approach is same as the proposed method except focusing only on articles.

- Evaluation metrics: Precision, Recall, F1-score

- Results

(a) Optimistic.

Method	Precision	Recall	F1-measure
Baseline	0.34	0.81	0.48
Proposed	0.51	0.79	0.62

including “others” class

(b) Skeptical.

Method	Precision	Recall	F1-measure
Baseline	0.39	0.83	0.53
Proposed	0.37	0.83	0.51

excluding “others” class

Use Case

Paragraphs about council in Landscape ordinances

Ami			Shichigashuku		
#A	#P	Article Title (Content desc.)	#A	#P	Article Title (Content desc.)
23	1	Establishment	11	1	Council (Establishment)
24	1	Deliberation matters	11	2	Council (Deliberation matters)
25	1	Counsel	11	3	Council (Counsel)
26	1	Organization (#Committee)	12	1	Organization (#Committee)
26	2	Organization (Conditions)	12	3	Organization (Conditions)
26	3	Organization (Temporary com.)			
27	1	Term of service (Basic)	12	2	Organization (Term of service)
27	2	Term of service (Substitute)			
27	3	Term of service (Temporary com.)			
28	1	Chairperson (Election)			
28	2	Chairperson (Chairperson)			
28	3	Chairperson (Vice chairperson)			
29	1	Convention (Summons)			
29	2	Convention (Resolution)			
30	1	Section			

Conclusion

- Motivation: support for OR document drafters
 - Searching relevant OR documents
 - Discovering parts for imitation
- Approach: relevancy metrics
 - Matching ratio: content matching
 - Provision commonality: structural matching
- Result
 - 0.62 F1 score for finding matches
 - Useful use case