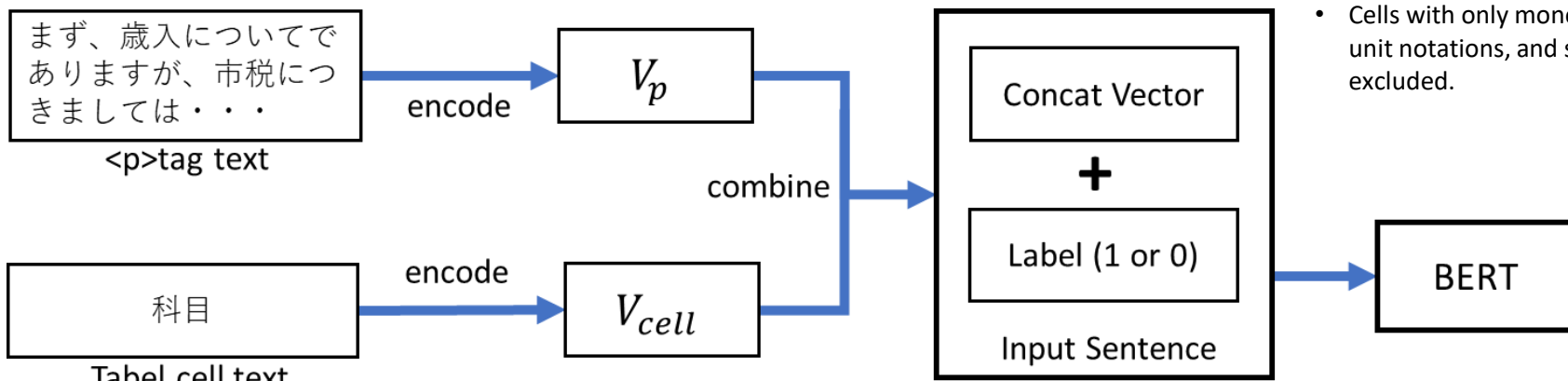


Method

The top 10 keywords were extracted, and the use of extracted words in place of text was also verified.



- We thought that related tables could be found by focusing on the cells of the table.
- Cells with only monetary amounts, unit notations, and symbols are excluded.

Table cells are extracted in **three ways** and combined with <p>tag.

<p>tag and cell-by-cell learning and inference. (Method 1)

科目	本年度		前年度		前年度	
	予算額	構成比	当初予算	構成比	予算額	構成比
	千円	%	千円	%	千円	%
1港湾整備事業費	244,465	40.7	244,465	40.7	244,465	40.7
2公債費	355,634	59.3	355,634	59.3	355,634	59.3
3予備費	100	0.0	100	0.0	-	-
歳出合計	600,197	100.0	584,675	100.0	15,522	2.7

Remove words that occur frequently. (Method 2)

<p>tag and column-wise learning and inference. (Method 3)

Method 3 :

- Join column by column using delimiters.
- Use "[SEP]" and "(comma)" as delimiters.

Inference :

- Look at the table one cell at a time.
- If the inference result of combining a cell and a p tag outputs that any one of them is related ("1"), it is assumed to be a related table.

とりわけ、市税などの歳入動向がコロナ禍により不透明な現状においては・・・

<p> tag text

Table1:	results
[
"科目",	0
"当初予算額",	0
"1市税",	1
"2地方譲与税",	
.	
.	
.	
]	
Table2	

Related Table

Cell word list

Result

Method 1	Method 2	Method 3 ([SEP])	Method 3 (,)	Keyword extraction ([SEP])	Keyword extraction (,)
1.43%	21.36%	27.23%	36.66%	11.63%	24.99%

Result :

- The best accuracy was obtained when "(comma)" was used as the delimiter in Method 3.

Consideration

Tables not linked by the proposed method. In the Gold data, it is linked with many <p> tags.

	会計別	本年度 予算額	前年度 当初予算額	比較	伸び率
		千円	千円	千円	%
	一般会計	56,639,525	55,047,861	1,591,664	2.9
	港湾整備事業	584,675	518,769	65,906	12.7
	青果物卸売市場事業	45,682	41,122	4,560	11.1
特	水産物卸売市場事業	35,119	37,164	△2,045	△5.5
	国民健康保険事業	18,407,404	18,472,977	△65,573	△0.4

Problem:

- Our methods could not link many tables. → Our methods does not handle tables with many rows well.
- Distributed Gold data may not be well developed. → The budget tables, such as the one on the left, tend to have links, but the detailed budget tables, such as the one below, tend to have no links in the Gold data.

Tables that were linked together by the proposed method. Not linked to the Gold data.

交付税	予算科目	本年度予算額	前年度予算額	比較
		千円	千円	千円
	11 地方交付税	16,581,000	15,340,000	1,241,000
	1地方交付税	16,581,000	15,340,000	1,241,000
	1地方交付税	16,581,000	15,340,000	1,241,000