WASEDA University

Smart SE

スマートエスイー 修了制作

ナレッジグラフを利用したSystematic Review作業支援

構造化された論文からのResearch Question回答に必要な部分構造の抽出

富士通株式会社　吉田裕輔　yoshida_yusuke@fujitsu.com

着目した問題点

Systematic Review(以降、SR)における、精査対象論文に対するResearch Question(以降、RQ)回答のための精査作業には、以下の問題のため、SRの品質を落とすリスクが高い。

- 精読必要範囲の絞り込みにくさ
- 絞り込んだ範囲からRQ観点に沿った要約/整理することに対する精読の難しさ

実現機能と価値仮説

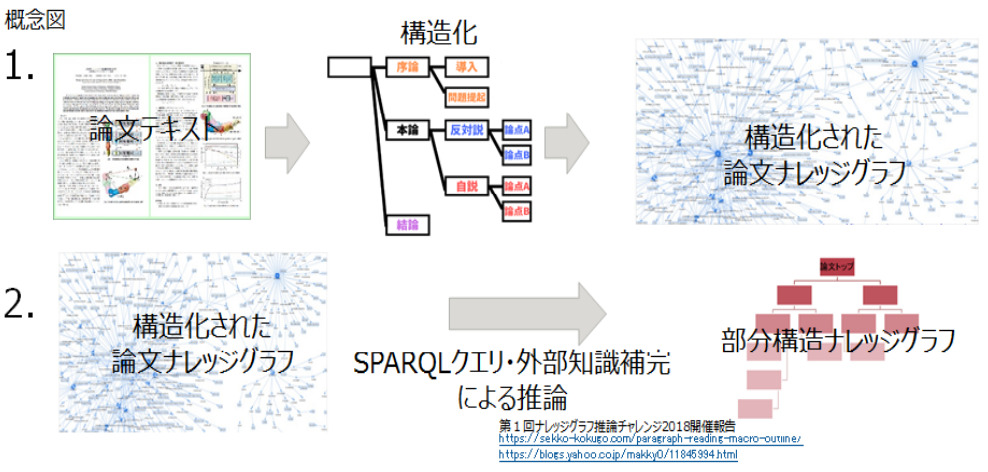
「論文の中からResearch Question回答に必要な箇所のみを抽出し、それを構造化した文章を提示する」ことで、SR実施者に以下の価値を提供する。

- 全文を探索することなく、必要な箇所のみを精読できることによる手間の軽減
- 文章構造を把握しながら精読することによる論文内容理解の効率化

アプローチ

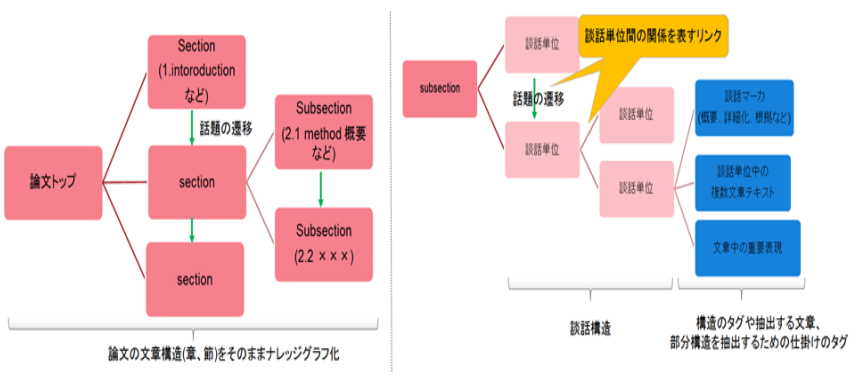
1. 論文を構造化し、それをナレッジグラフで表現する
2. 構造化された論文ナレッジグラフに対して、RQ回答のために必要である適切な範囲の部分構造ナレッジグラフを、SPARQLクエリ・外部知識補完による検索を利用して抽出する

概念図



ナレッジグラフ化設計

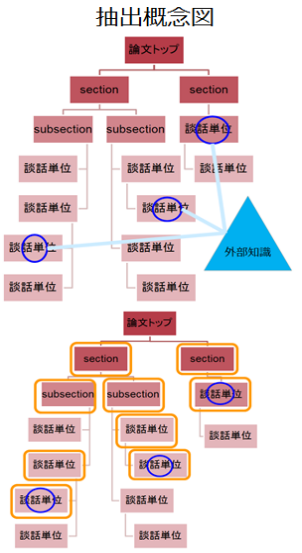
1. 談話構造化により、章/節以下の文章をさらに詳細な単位(=談話単位)に分割する
2. 談話単位をノードとして、そこに抽出のため情報(右下図:青四角図)のつながりを付与する
3. 上記結果と論文の文章構造(章/節)と合わせて、階層的に構造化する



部分構造ナレッジグラフ抽出

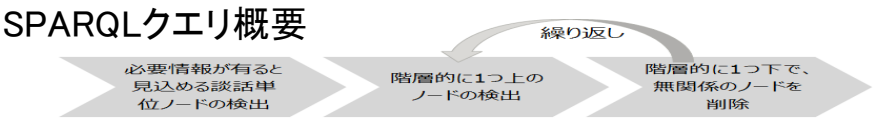
1. 外部知識補完により、“要約/整理に必要な情報”と談話単位ノードにつながりを付与し、そのつながりの先である談話単位ノードを検出する(右図:青丸)
2. 適切な範囲の部分構造を「検出した談話単位ノードをすべて通る最小限のグラフ」と定義し、これをSPAQRLクエリによって抽出する(右図:オレンジ四角)

抽出概念図



プロトタイプ

SPARQLクエリ概要

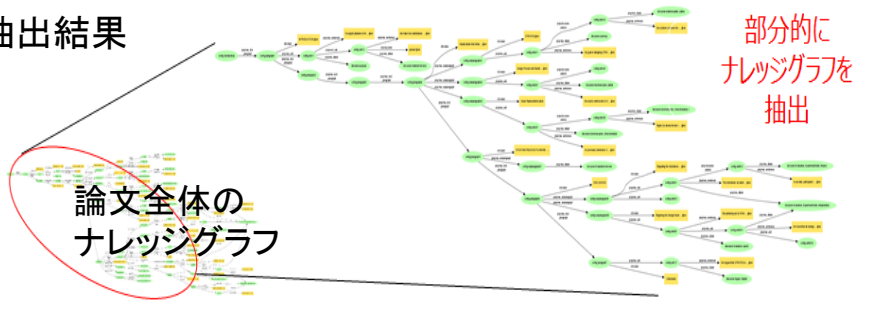



```
SELECT ?s ?o
WHERE
{
{?s clue:keyword ?o .
FILTER (?o = "goal"@en || ?o = "present"@en)
}
UNION
{?s clue:keyword_org ?o .
FILTER (?o = "visual notations"@en)
}
}
```

```
SELECT ?s
WHERE
{
{?s prop:has_unit ?o.
FILTER (?o = config:unit131 )
}
UNION
{?s prop:has_subsection ?o .
FILTER (?o = config:unit131 )
}
}
```

```
DELETE DATA
{ GRAPH :"グラフIRI"
{ config:subsection13 prop:has_unit
                config:subsection132 . } }
```

抽出結果



スマートエスイー　スマートシステム＆サービス技術の産学連携イノベーティブ人材育成