

学術論文の意味検索の為の Domain-Ontology の設計

金奎煥(Kim GyuHwan emm1996@gmail.com),

南泳準(Nam YoungJoon), 李汀鎬(Lee JungHo)(韓國中央大学大学院 文献情報学科)

1. 序論

1.1 研究の必要性と目的

学術論文の量が増えている今、利用者たちは利用目的に適切な学術情報を検索することに多い時間と努力が必要になっている。従って、学術論文の検索システムに最も重要な機能は利用者の検索意図を反映した正確な検索結果だけを提供することである。しかし、大抵の検索システムは著者キーワードと題目、抄録の文字列と利用者の質疑語の間のシミラリティーを考慮した単純な Keywords matching 検索技法を使用している。その検索技法は利用者の検索意図と関連性がない情報も検索され、利用者は検索結果を再検索する不便なことがある。

最近 XML 基盤の Meta 情報の表現技術である PDF/PDF(s)、OWL 等を利用した Ontology を情報検索に適用する研究が活発に行われている。本研究は利用者の検索意図に符合する検索結果を提供する方法として、学術論文の題目に出現する keywords に意味情報をつける Domain-Ontology を設計した。

1.2 研究の内容と方法

実際学術論文の題目に含まれた keywords 間の概念構造、つまり ontology を modeling した。具体的な研究内容と方法は、

1) 韓国の文献情報学分野の論文題目に現

れた構文及び意味的な構造を分析した。それから論文題目の keywords の役割概念と概念関係を導出した。

2) 最終的に 10 の概念 class と 9 の関係属性を抽出し ontology を設計した。

2. 論文題目の構成要素及び意味解析

学術論文の題目には一定した構文及び意味構造の pattern が現れている。こうした pattern では論文題目に含まれた keywords は一定した役割概念を持っていることを分析することができる。役割概念とは学術論文に著者が表現したい学術的概念である。

論文題目は一般的に名詞及び名詞句と機能語 (functional words) で構成されている。名詞及び名詞句はその論文の核心 idea を伝達する重要な用語及び keywords の役割をし、助詞、前置詞、接触詞のような機能語は keywords の間に位置し、keywords の関係す定義する Clue-word の役割をする。例えば、「機械学習を利用したインターネット学術文書の自動分類」で、Clue-word は論文題目から keywords の役割を示し、keywords 間の意味関係を示す役割 (Syntactic Role) をする。「インターネット学術文書」は「研究対象」を、「を利用した」という Clue-word を通じて「機械学習」は「使用道具・手段」の役割概念を持つ。つまり、「インターネット学術文書」は「を利用した」という関係属性を通

じて「機械学習」と意味的に連結できる。そして、「機械学習」は「の」Clue-wordを通じて「自動分類」と「研究対象と関連な行為」という役割概念と繋がる。

3. 論文題目構文及び意味構造分析を通じたOntologyの設計

3.1 分析対象及び分析方法

本研究では韓国の文献情報分野、三つの学会誌を対象とし、2000年から2004年まで収録された学術論文から805の学術論文の題目を収集・分析している。

3.2 論文題目の構文及び意味構造の分析

1) keywordsとClue-wordの結合pattern

分析結果、分析対象の論文題目からkeywordとClue-wordの結合patternが9つあることがわかった。

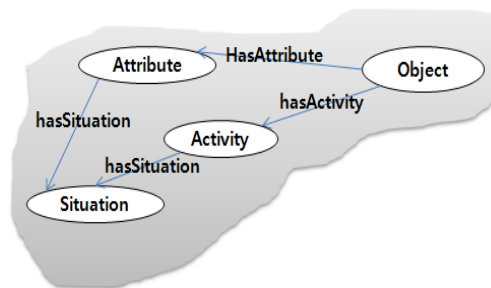
- (1) 「K1+の+K2」、「K1+K2」
- (2) 「K1+に+K2」
- (3) 「K1+を利用した+K2」
- (4) 「K1+に基盤した+K2」
- (5) 「K1+の為の+K2」
- (6) 「K1+を中心に+K2」
- (7) 「K1+と+K2」
- (8) 「K1+的研究」
- (9) 「K1+が+K2+の+K3に与える影響」

※ K1, K2は論文題目に使われたkeywordsを意味する。

2) 論文題目の構文及び意味構造

実際、論文題目の構文構造は「keywordsとClue-wordとの結合pattern」の中の一つ

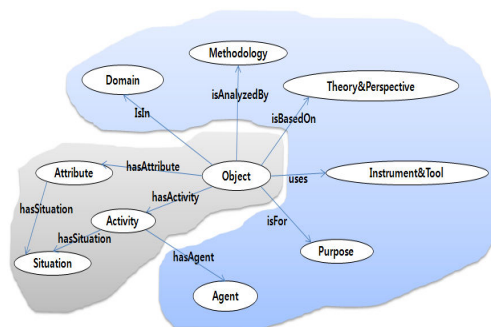
及び2~3つの結合patternが調査された形をもっていた。本研究では1つの基本の意味構造と6つの拡張意味構造が導出された。基本構文及び意味構造は<図1>のようである。



<図1 基本構文及び意味構造>

論文題目の基本構文及び意味構造は研究対象 (object) に関連行為 (activity)、属性 (attribute)、状況 (situation) の概念 class とその関係属性である「has attribute」、「has activity」、「has situation」で構成されている。

拡張構文及び意味構造は<図2>のように、研究対象を中心とした基本構文及び意味構造を中心として拡張した構文及び意味構造を提示している。



<図2 拡張構文及び意味構造>

以上の論文題目の定形化した構文及び意味構造を利用して、学術論文に含まれた概

念の組織を次のように表現することができる。

「学術論文 Article は研究目的 (purpose) を達成する為に研究範囲 (domain) 内の研究理論 (theory) 及び研究観点 (perspective) に基盤して、研究対象 (object) と関連属性 (attribute) に対して、関連作用 (activity) に対し、及び原因又は行為主体 (agent) の影響を受けることに対し、研究手段 (instruments) と道具 (tool) を利用したり、研究方法論 (methodology) を使用したもの。」

学術論文は本研究から見出した概念間の意味構造を通じて学術的な意味解析が可能である。これを情報検索に適用する場合、学術論文を検索する利用者は様々なMeta情報が付けられたkeywordsを対象と自分の情報要求に最も接近性が近い学術論文を検索する可能性が高まる。

3.3 ontologyの設計

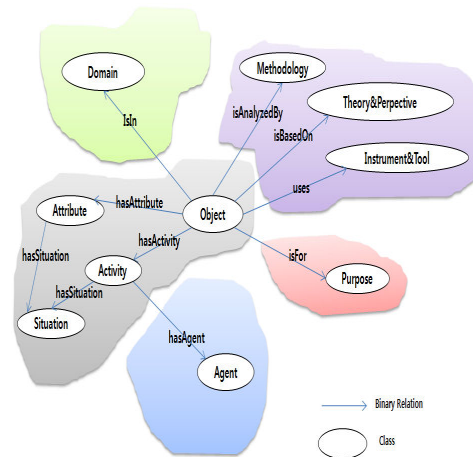
Ontologyを構成するclassと関係属性を整理すれば<表1>になる。

classes	Properties
- object	- has object
- purpose	- is for
- instrument & tool	- uses
- basic theory & perspective	- is based on
- Methodology	- is analyzed by
- domain	- is in

- activity	- has activity
- attribute	- has attribute
- situation	- has situation
- agent	- has agent
- title	

<表1 classと関係属性>

<表1>の論文題目のkeywordsの役割概念と関係属性を図式化すると<図3>になる。



<図3 keywordsの役割概念と関係属性>

次は論文題目「client sever環境で韓国語 text の検索の為の best match algorithms の具現」に、keywordsにmeta意味情報を付けた結果である。

```
<KRO:Title rdf:ID="client sever 環境で韓国語 text の検索の為の best match algorithms の具現">
<KRO:hasObject>
<KRO:Object rdf:ID=" best match algorithms ">
<KRO:hasActivity>
<KRO:Activity rdf:ID=" 具現">
</KRO:hasActivity>
<KRO:isIn>
<KRO:Domain rdf:ID="client sever 環境 "/>
</KRO:isIn>
<KRO:isFor>
```

```
<KRO:Purpose rdf:ID="韓国語 text の検索"/>
</KRO:isFor>
</KRO:Object>
</KRO:hasObject>
</KRO:Title>
```

※KROはName spaceの接頭語

4. 情報検索の活用方案

学術論文は本研究から見出した概念間の連結構造を通じて様々な意味解析が可能である。これを情報検索に適用する場合、学術論文を検索する利用者は様々なmeta情報が付けられたkeywordsで自分が意図した正確な対象論文を検索することができる。それから、その検索結果の集合の数も最適にすることができる。

つまり、単純なkeywords基盤の検索よりmeta情報(semantic information)を利用した検索結果は部分集合である。最適な場合、利用者が意図する正確な検索結果を提示することが可能で、最悪の場合でも既存のkeywords基盤の検索と同じ結果がでることである。

5. 結論及び提言

本研究は利用者の検索意図を識別し、正確な検索結果を提供する為に、論文題目の構文及び意味構造を分析し、論文題目に含まれたkeywordsの役割情報をsemantic情報で構成して論文検索に利用する方法を提案した。

本研究は韓国の文献情報学の分野に限定して、論文題目に現われた構文及び意味構造を基盤として設計したprototypeの

domain-ontologyである。今後、もっと様々な分野で適用できるように先のdomain-ontologyの拡張及び精巧化した後続の研究が必要である。本研究の研究結果はこれからの研究の為に基礎ontologyとしての役割を期待することができる。

参考文献

Kim YoungMin, Lee SangJoon 2003. Semantic based Research-Paper Searching System. 「Korean Society For Internet Information」, 4(3). pp. 15-22.

Song InSeok. 2008. A Study on Ontology-based Keywords Structuring for Efficient Information Retrieval. 「Journal of Information Management」, 34(9). pp. 121-154.

Sano, H., & Fujiwara, Y. 1993. "Syntactic and Semantic Structure Analysis of Article Titles in Analytical Chemistry". Journal of Information Science, 19(2). pp. 119-124.

Forsbom, E. 2005. Rhetorical Structure Theory in Natural Language Generation.

Gross, A. 1996. The Rhetoric of Science: Harvard University Press.