

構文解析による発想支援アーキテクチャの検討 Study of Creativity Support Architecture by Syntactic Analysis

楨 俊孝[†] 若原 俊彦[†]
Toshitaka Maki Toshihiko Wakahara

1. はじめに

本研究では、ネットワーク・アーキテクチャを模擬した発想支援アーキテクチャを提案し、これを実現する発想支援ツールの検討を行う。発想支援の手法には、従来から様々なものがあり、代表的なものとして KJ 法やブレインストーミング、マインドマップなどがある。また、コンピュータを利用した発想支援ツールの研究開発も多く行われており[1]、時間や場所などの制約のもとで効率的に発想を行うことができるように工夫されている。

2. 関連研究

関連研究としては連想検索エンジン[2]などがあり、これに基づく発想支援ツールがある。本研究では、新たに構文解析システムを基盤技術として発想支援アーキテクチャ (CSA : Creativity Support Architecture) を構築し、これに基づいた支援ツールを実現する。具体的には、ネット上で流行っている文章の中から、発生頻度の高い単語を抽出して、構文解析から係り受けの単語のツリーの構成を他のドキュメントの文章中の単語にリンクさせ、関連する単語を抽出して発想支援に利用するものである。

3. 発想支援アーキテクチャ (CSA) の提案

CSA は、KNP[3]などの構文解析ツールによりネット上の文章を抽出して構文解析し、ネットワーク・アーキテクチャを模擬して発想支援ツールを構築する。3.1 節ではこのネットワーク・アーキテクチャの代表である OSI 参照モデルにおけるレイヤ 2 とレイヤ 4 について述べ、3.2 節では発想支援アーキテクチャを構築するための諸条件について述べる。さらに、3.3 節では CSA の各機能、3.4 節では CSA によるテキスト分析について述べる。

3.1 OSI 参照モデルによる通信機能

本研究では、OSI 参照モデルにおけるレイヤ 2 (データリンク層) とレイヤ 4 (トランスポート層) の機能を模擬して発想支援アーキテクチャを構築する。このアーキテクチャでレイヤ 2 は隣接しているノード間で通信を行い、レイヤ 4 はネットワークの端から端までのリンク管理を行う。

3.2 CSA におけるルール

本研究で提案している発想支援は、キーワード間の関係性をユーザに示し、ユーザが連想できないような知識構造をユーザに提示するものである。これを実現するために以下に示すルールを適用する。

- ① キーワードは一般形の名詞と形容詞である。
- ② 文における主語は目的キーワードである。
- ③ 文における述語は目的キーワードの説明である。

- ④ 動詞は接続詞として扱う。
- ⑤ 経路候補キーワードは、第 1 層キーワードに関係するものであり、エッジ接続数が大きいものである。

3.3 CSA の機能

発想支援アーキテクチャは、図 1 に示す 3 つのレイヤから構築される。以下では、それぞれのレイヤについて説明する。

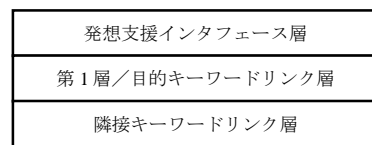


図 1 発想支援アーキテクチャの構成

3.3.1 隣接キーワードリンク層 (レイヤ 1)

このレイヤでは、OSI 参照モデルにおけるデータリンク層のような機能を実現する。隣接するキーワードは単方向リストで管理しユーザが特定のキーワードを選択したときに次にキーワードを提示できるようにする。また、隣接キーワード間の繋がりを共起頻度によって評価しているため、よく利用されているキーワードを優先的に選択することができる。

3.3.2 第 1 層/目的キーワードリンク層 (レイヤ 2)

このレイヤでは、OSI 参照モデルにおけるトランスポート層のような機能を実現する。第 1 層キーワードと目的キーワードを管理し、隣接キーワードの提示をサポートする。このレイヤによってキーワードの意味が発散することを防ぎ、キーワード全体で 1 つの意味を成す経路を示すことができ、レイヤ 1 と同様に出現頻度によって評価をしている。

3.3.3 発想支援インタフェース層 (レイヤ 3)

このレイヤでは、発想支援を行うためのインタフェースを提供する。具体的には、レイヤ 1 とレイヤ 2 で管理しているキーワードをインタフェースのボタンとして表示し、ユーザがそのボタンを押すことによって関連するキーワードをボタン表示するものである。なお、ボタンの表示順序は下記の (式 1) により求め、重みの高い順序で表示を行う。つまり、媒介エッジのルートが多い経路を選択できるようにし、効果的な発想支援を図る。

$$W_{n,e,i} = \sum_j w_{n,i,j} + \sum_k w_{e,i,k} \cdots \quad (\text{式 1})$$

$w_{n,i,j}$ は発想支援アーキテクチャにおけるレイヤ 1 の隣接キーワード毎の頻度を表しており、 $w_{e,i,k}$ はレイヤ 2 の第 1 層/目的キーワード毎の頻度を表している。この (式 1) に従って重み $W_{n,e,i}$ を導出し、効果的なキーワード群を導き出す。なお、 $w_{n,i,j}$ はレイヤ 2 において有効と判断された隣接キーワードである。

[†] 福岡工業大学 大学院 情報通信工学専攻
Graduate School, Fukuoka Institute of Technology

3.4 CSAによるテキスト解析

本研究では、KNPにより論文を構文解析した結果をCSAによって評価をして発想支援を行う。図2に処理手順の概略図を示す。

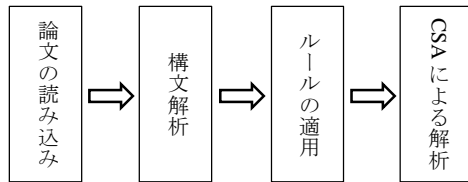


図2 CSAによる処理手順

具体例を以下に示す。

① 論文の読み込み

下記のサンプル文を読み込んだとする。

特別研究員制度は、我が国トップクラスの優れた若手研究者に対して、自由な発想のもとに主体的に研究課題等を選びながら研究に専念する機会を与え、我が国の学術研究を担う創造性とんだ研究者の養成・確保を図る制度です。

② 構文解析とルールの適用

サンプル文を構文解析し、3.2節で示したルールを適用することで図3のようなキーワードネットワークが構築される。

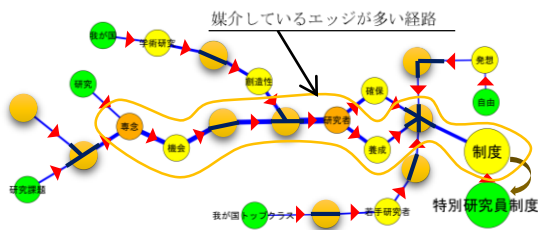


図3 キーワードネットワークの構築例

図3を見て分かるように目的キーワード「特別研究員制度」への経路は複数ある。この場合は、媒介エッジが多い経路を、利用頻度の高い一般的なキーワードの組み合わせと仮定して正規ルートとする。これは、(式1)に示した重み評価によって決定することができる。このキーワードネットワークの情報は図4に示すデータベースに格納し、レイヤ3で利用できるようにしている。

Keyword テーブル	<table border="1"><tr><th>kevid</th><th>keyword</th></tr></table>	kevid	keyword		
kevid	keyword				
NN テーブル	<table border="1"><tr><th>nid</th><th>keyid_b</th><th>keyid_a</th><th>freq</th></tr></table>	nid	keyid_b	keyid_a	freq
nid	keyid_b	keyid_a	freq		
EE テーブル 1	<table border="1"><tr><th>eid1</th><th>keyid_f</th><th>keyid_e</th><th>freq</th></tr></table>	eid1	keyid_f	keyid_e	freq
eid1	keyid_f	keyid_e	freq		
EE テーブル 2	<table border="1"><tr><th>eid2</th><th>keyid</th><th>eid1</th></tr></table>	eid2	keyid	eid1	
eid2	keyid	eid1			

図4 CSAにおけるデータベース構造

Keyword テーブルには構文解析してルールを適用した名詞と形容詞を格納している。レイヤ1のNN (Neighboring Nodes) テーブルでは、隣接キーワードを単方向リストで関連付け、データリンク層のような機能を実現している。また、レイヤ2の2つのEE (End to End) テーブルでは、第1層キーワードと目的キーワードを単方向リストでキーワードの1つ1つに関連付け、トランスポート層のような

機能を実現している。これらの4つのデータベースをCSAで制御し、効果的な発想支援を図っている。

4. CSAの実例

今回は、「教育」に関する論文150件を構文解析し、CSAに基づいた実例を図5に示す。図5は、Webブラウザに表示した発想支援ツールの画面である。



(a) 第1層の画面

(b) 第4層の画面

図5 CSAの実演

図5(a)に示す第1層の画面は、「教育」に関する論文を解析して自動的に構築されたものである。第2層、第3層とキーワードを選択していくと第4層では図5(b)の画面が出力される。このようにCSAによって自動的にカテゴリ階層を作ることができる。

5. まとめ

本研究では、構文解析システムを基盤技術として新たに発想支援アーキテクチャ(CSA)を構築し、これに基づいた発想支援ツールを実現した。文章の中から、発生頻度の高い単語を抽出し、構文解析から係り受けの単語のツリーの構成を他のドキュメントの文章中の単語にリンクさせ、関連する単語を抽出して発想支援に利用している。CSAは3つのレイヤから構築し、ユーザが選択したキーワード全体で1つの意味を持つようにしている。CSAの実例により自動的にカテゴリ階層が構築され、CSAの有効性を確認することができた。今後の課題は、より多くの論文を解析してボキャブラリーを増やし、また、統計的にCSAを分析して有効性を実証することである。

参考文献

- [1] 発想支援ツール
<http://ideatool.jp/index.php?%E7%99%BA%E6%83%B3%E6%94%AF%E6%8F%B4%E3%83%84%E3%83%BC%E3%83%AB>
- [2] 連想検索エンジン reflexa <http://labs.preferred.jp/reflexa/>
- [3] 黒橋 禎夫, 河原 大輔, “京都大学自然言語処理ツール”, 情報処理学会研究報告, Vol.53 (2000).