

レシピフローグラフへの簡潔データ構造の適用

並木 拓哉

要旨

レシピフローグラフとは、Mori, Yamakata らによって提案された料理レシピの抽象表現であり、各ノードをレシピ用語、各エッジをノード間の関係で表現する根付き有向無閉路グラフである。レシピフローグラフは、グラフ表現を利用した料理の理解に加え、固有表現認識や述語項構造解析、共参照解析といったレシピテキストの理解に必要とされる諸技術への貢献が期待されている。クックパッドでは、2018年11月末に全世界合計で投稿レシピ数が500万件を突破し、レシピフローグラフを用いた料理理解に関する研究の重要性がますます重要となるが、その一方で、データの増加に伴う処理速度の低下とデータ記憶領域の圧迫が一つの問題となると予想される。

簡潔データ構造とは、情報理論的下界に近い領域量だけを使いつつ、検索等の処理を行う際にはあたかも非圧縮のデータに対してアクセスしているように扱えるデータ構造である。近年の情報サービスやビッグデータの増加に伴い、簡潔データ構造の持つ高い空間効率は注目を集めている。その一種である順序木を表現する LOUDS は、非常に少ないメモリで木を表現しながら、高速に検索を行うことができる。これまでに、LOUDS のパフォーマンス向上を図る LOUDS++ や、LOUDS におけるビットベクトルに付与する索引の最適化、LOUDS を応用したグラフの簡潔表現である GLOUDS、ファイルシステムの簡潔表現である FLOUDS など、LOUDS に関連した様々な応用が提案されてきた。

本研究では、LOUDS と、GLOUDS を用い、レシピフローグラフデータの簡潔な表現を図った。また、実際のデータを用いてそれらを処理速度と空間効率の観点から比較し、その実用性を評価した。