

研究タイトル:

自然言語処理を用いた容量削減技術の研究



氏名:	横山ひなた / Hinata YOKOYAMA	E-mail:	yokoyama-h@di.kagawa-nct.ac.jp
職名:	助教	学位:	修士(工学)
所属学会・協会:	情報処理学会		
キーワード:	情報ストレージシステム, 自然言語処理, ストレージ容量削減技術		
技術相談 提供可能技術:	・情報ストレージシステムに関する研究相談		

研究内容:

背景

近年、生成されるデータ量は年々増加している。生成されるデータのうち、特に重要なデータはバックアップストレージに保存される。生成されるデータ量の増加に伴いバックアップされるデータ量も増加している。

バックアップストレージは、ストレージ容量とデータ保存時の負荷が重視される。ストレージ容量を最適化する一般的な技術として、圧縮や重複排除が用いられる。これらの技術は、冗長なデータを検出して、符号化および削除することでストレージ容量を削減する。しかし、冗長なデータを検出するために、ファイル全体を読み込む必要があるなど、負荷が高くなる。

課題

データ容量削減技術における負荷を抑制するために、読み込むファイル数を絞り込む。絞り込み方式の従来方法として、ファイルサイズや拡張子がある。しかし、データ量増加によって同一サイズや拡張子のファイルが増加したことで、絞り込み効果が低下している。そのため、新しい絞り込み方式が必要となる。

提案

本研究は新しい絞り込み方式として、自然言語処理を用いて算出した類似度を提案する。ファイル名やディレクトリ名に対して自然言語処理を用いて類似度を算出することで、冗長データを持つファイルを絞り込む。本システムは図1に示すように、絞り込み処理とデータ量削減部に分けられる。絞り込み処理では、従来手法に自然言語処理を追加する。絞り込み処理で作成した類似度リストからデータ量削減部がファイルシステムに対して容量削減を行う。

本研究は、類似度の閾値設定にモデル式を作成する。モデル式を用いることで適切なファイル類似度の閾値とデータ量削減技術が選択できる。

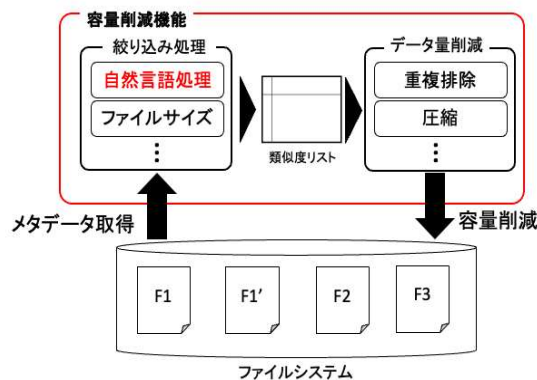


図1 自然言語処理を用いた容量削減システムの概要

提供可能な設備・機器:

名称・型番(メーカー)

名称・型番(メーカー)	