

自己評価報告書

平成23年 3月20日現在

機関番号：25403
研究種目：基盤研究（C）
研究期間：平成20年度～平成24年度
課題番号：20500137
研究課題名（和文） 知識創造支援型データベースシステムの構成法と効率化に関する研究
研究課題名（英文） Method for Organizing Knowledge Creating Database and Achieving Efficient Processing
研究代表者
北上 始 (KITAKAMI HAJIME)
広島市立大学大学院情報科学研究科・教授
研究者番号：50234240

研究分野：総合領域

科研費の分科・細目：情報学・知能情報学

キーワード：テキストマイニング、配列マイニング、索引構造、正規表現、並列処理

1. 研究計画の概要

本研究では、人間の知識創造過程が記憶・演繹・帰納・発想から成るという点に着目し、資産価値や知識資産の創造の支援をめざした知識創造型データベースシステムの研究を行う。このために、(1)知識創造支援機能の構成法の研究、(2)知識創造支援処理の効率化の研究、(3)知識創造過程におけるコミュニケーションの成立を支援する方法の研究、(4)知識創造支援機能を応用する方法の研究を行う。

2. 研究の進捗状況

前述の4項目のそれぞれの進捗状況については、以下のとおりである。

(1) 知識創造支援機能の構成法

文字列データ（特に、分子配列データ）に含まれる曖昧性に注目し、文字列データに対する類似検索結果（関係オブジェクト）に対しては、規則性を見つけ出すための最小汎化処理の方法^[学会発表⑤]について、部分的ではあるが明らかにした。また、類似検索結果に対して、マルチプルアラインメントを予め適用することが最小汎化処理の精度向上につながる事が分かった^[学会発表①]。その他、頻出な類似パターン抽出結果(KDDオブジェクト)に対しても、マルチプルアラインメントの適用が、その後の類似パターン抽出において、精度向上に多少寄与することもわかった^[学会発表③]。

(2) 知識創造支援処理の効率化

文字列データの類似検索結果に対する最小汎化処理を効率的に行う方法^[雑誌論文①]を明らかにした。また、特定の空間データ（蛋白質立体構造データ）に対する高速類似検索のために有用なディスク版のサフィック

ス木を作成する方法^[学会発表②]について検討を行った。さらに、データは特定しないが、大規模なディスク版のサフィックス木の高速化のために、サフィックス木を並列化する方法^[学会発表④]についても検討を行った。

(3) 知識創造過程におけるコミュニケーションの成立を支援する方法
ブログ空間に限定し、コミュニケーションを支えるコミュニティを抽出する方法を検討した^[雑誌論文②,④]。また、調停グラフの効率的な視覚化の方法を検討した^[雑誌論文③]。

(4) 知識創造支援機能を応用する方法
産学連携によるマッチングセッションや広島市役所での展示会などでのポスター展示を初めとして、大学附属図書館との研究交流会などを通じ、図書推薦、Webマイニング、バイオインフォマティクス、テキスト処理などの分野への応用可能性についての検討を行った。

3. 現在までの達成度

(1) 知識創造支援機能の構成法

②おおむね順調に進展している

（理由）国際会議等で発表できる内容が出てきている。

(2) 知識創造支援処理の効率化

①当初の計画以上に進展している

（理由）国際会議等で発表できる内容が出てきている。また、知識創造支援処理の一部ではあるが、雑誌論文にその成果をだせるまでに至っている。

(3) 知識創造過程におけるコミュニケーションの成立を支援する方法

③やや遅れている。

（理由）コミュニケーションの視覚化という点では一部の課題が解決されているもの

の、残された課題が未だある。
(4) 知識創造支援機能を応用する方法
② おおむね順調に進展している
(理由) 産学連携イベントなどでの出展参加では、上記(1)~(3)についての研究内容やそれらの応用の方向性を見失っていないことを確認している。

4. 今後の研究の推進方策

- (1) 知識創造支援機能の構成法
空間データや文字列データを対象にした曖昧性処理を中心に、知識創造支援機能の研究を推進する。
- (2) 知識創造支援処理の効率化
索引構造や並列処理を中心に、スケーラビリティを意図した効率化の研究を推進する。
- (3) 知識創造過程におけるコミュニケーションの成立を支援する方法
CG技術などを積極的に導入し、コミュニケーションを視覚化する研究を推進する。
- (4) 知識創造支援機能を応用する方法
広島市立大学の社会連携センターを積極的に活用し、産学交流イベントに可能な限り参加し、研究の応用範囲について検討する機会を設ける。また、高校生を対象にした公開講座などに、研究成果を還元する。

5. 代表的な研究成果

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計4件)

- ① 田村 慶一, 木村 浩明, 荒木 康太郎, 北上 始: 段階的一般化法によるミスマッチクラスタを表現する最小汎化集合の効率の抽出, 電子情報通信学会和文論文誌D「データ工学特集号」, 電子情報通信学会, Vol. J93-D, No. 3, pp.189-202, 2010年3月, 査読有り.
- ② 高木 允, 田村 慶一, 森 康真, 北上 始: 密な部分構造抽出のための階層的凝集型クラスタリング手法, 日本データベース学会論文誌, Vol. 7 No. 1, pp. 275-280, 2008年6月, 査読有り.
- ③ 田村 慶一, 森 康真, 北上 始: Extremal Optimizationによる調停グラフの交差数減少, 情報処理学会論文誌: 数理モデルと応用, Vol. 4, No. SIG4 (TOM20), pp. 105-116, 2008年3月, 査読有り.
- ④ 高木 允, 森 康真, 田村 慶一, 北上 始: ブログユーザ空間からの重複を許した頻出コミュニティ抽出法, 情報処理学会論文誌: 数理モデルと応用, Vol. 49, No. SIG4 (TOM20), pp. 93-104, 2008年3月, 査読有り.

[学会発表] (計5件)

- ① Kazuya MIYAHARA, Hajime KITAKAMI,

Yoshifumi TAKAHASHI, Keiichi TAMURA and Susumu KUROKI: Mining Minimum Generalized Set Based on Multiple Alignments from Mismatch Cluster, Proceedings of The 2010 International Conference on Bioinformatics and Computational Biology (BIOCOMP'10), CSREA Press, Vol. I, pp.35-41, 2010, 査読有り.

- ② Yoshifumi TAKAHASHI, Susumu KUROKI, and Hajime KITAKAMI: Efficient Query Processing in Protein Structure Databases, Proceedings of the 2nd International Workshop with Mentors on Databases, Web and Information Management for Young Researchers (iDB Workshop 2010), pp.47-55, 2nd - 4th August, 2010, 査読有り.
- ③ 河野 修久, 田村 慶一, 森 康真, 北上 始: ギブスサンプリングとアラインメント処理に基づく類似部分配列の抽出方式, 情報処理学会・数理モデル化と問題解決 (MPS) およびバイオ情報学 (BIO), Vol.2009-MPS-76 No.46 および Vol.2009-BIO-19 No.46, 8 pages, 2009年12月, 査読無し.
- ④ Yusuke SAWADA, Keiichi TAMURA, Kotaro ARAKI, Makoto TAKAKI, and Hajime KITAKAMI: Parallel Construction Method of a Disk-Based Suffix Tree on a PC Cluster, Proceedings of the 2008 International Conference on Parallel & Distributed Processing Techniques & Applications (PDPTA'08), Vol. II, pp.797-803, Las Vegas in USA, July 14-17 in 2008, 査読有り.
- ⑤ Hiroaki KIMURA, Hajime KITAKAMI, Kotaro ARAKI, and Keiichi TAMURA: A Stepwise Generalization Method for Extracting Minimum Generalized Set from Mismatch Cluster, Proceedings of the 2008 International Conference on Bioinformatics and Computational Biology (BIOCOMP'08), Vol. II, pp.998-1004, Las Vegas in USA, July 14-17 in 2008, 査読有り.

[図書] (計0件)

[産業財産権]

○出願状況 (計0件)

○取得状況 (計0件)

[その他]

特になし