

ローカルデータ・公開データを統合利用可能なコンテキスト検索エンジンの提案

Proposal of Context Search Engine for Integrated Use of Open and Private Data

岡久太一 高間康史
Taichi Okahisa Yasufumi Takama

首都大学東京大学院システムデザイン研究科
Graduate School of System Design, Tokyo Metropolitan University

This paper aims to extend the context search engine so that it can retrieve both of open and private data. The context search engine has been developed for answering trend-related queries. While it is designed to store and retrieve publicly available data (open data), it is supposed that users often have data that can't be disclosed to the outside due to problems such as privacy protection. Making it possible to access both of such local data and open data in integrated manner would lead to effective utilization of data such as value creation from data. In order to examine the utility value of local data without sharing its contents, this paper introduces the concept of Data Jacket into the context search engine. Experimental results show its effectiveness.

1. はじめに

本稿ではローカルデータと公開データを統合利用可能なコンテキスト検索エンジンを提案する。「動向に関する問い」というドメインに依存しないタスクに特化することで、より高度な基本検索機能を提供するコンテキスト検索エンジンが研究されている [高間 15]。従来のコンテキスト検索エンジンは公開可能なデータをサーバに集約したものを検索対象としており、プライバシー保護などの問題で外部に公開できないようなデータを利用することができない。しかし、ローカルデータを外部公開データと組み合わせることは、データからの新たな価値創造など、その有効活用につながると考える。そこで本稿では、ローカルデータと公開データの両者を統合して検索可能なシステムを提案する。具体的には、データの概要情報のみを記載した Data Jacket [大澤 14] を共有することによって、非公開データ自体を外部に公開することなく、データの利用価値を検討することを可能とする。プロトタイプシステムを実装し、ユーザ実験によって提案手法の有効性を示す。

2. 関連研究

2.1 コンテキスト検索エンジン

コンテキスト検索エンジン [高間 15] とは、「動向に関する問い」というドメインに依存しないタスクに特化した、既存検索エンジンよりも高度な検索が可能な検索エンジンである。Web上に公開されている動向情報を収集し、特徴的変動を抽出することで、動向情報の検索を可能とする。「検索期間」「アイテム名」「特徴的変動」「検索結果の出力タイプ」「単位」「統計データ名」がクエリとして入力可能であり、以下の3タイプの基本検索機能と組み合わせることで高度な検索が可能となる。

1. 指定したアイテムに関する動向が特徴的変動を示した期間の検索
2. 指定した期間に特徴的変動を示したアイテム・動向の検索

3. 指定したアイテムに関する動向が特徴的変動を示した期間に [同様の／対照的な] 変動を示したアイテム・動向の検索

2.2 Data Jacket Store

Data Jacket [大澤 14] とは、データの内容を説明する概要情報を記載したものである。これを共有することで、プライバシー保護などの問題で外部に公開することができないようなデータの利用価値を検討することが可能となる。早矢仕ら [早矢仕 16] は、過去に活用された Data Jacket から、自分以外のユーザが考案したデータの活用方法を発見可能とするために、データ利活用知識モデルを提案している。これに基づき、データ利活用方法検討を支援する検索システム Data Jacket Store を開発している。

3. 提案インタフェース

3.1 ローカルデータ・公開データの統合検索機能

従来のコンテキスト検索エンジンは、Web上に公開されている動向情報を収集し、サーバに集約したものを検索対象としていた。収集した動向情報から特徴的変動を抽出してデータベースに格納することで、コンテキスト検索エンジンで検索可能となる。この処理をユーザがローカルに保持している時系列データを対象に行い、検索可能にする。提案システムでは、外部サーバに登録されている公開データを検索可能にするために、コンテキスト検索エンジンが提供する API 機能を利用し、ローカルデータ・公開データを選択または統合して検索可能とする。

3.2 Data Jacket の導入

コンテキスト検索エンジンに Data Jacket を導入することで、ローカルデータと外部公開データの統合利用に貢献することが期待できる。データ利活用知識モデルによるアプローチ [早矢仕 16] を採用し、Data Jacket 検索システムを導入することで、過去に自分以外のユーザが発見した知見から、新たな知見の発見が可能となることが期待できる。コンテキスト検索エンジンでは、動向情報の変動に関して発見した知見を共有することが重要との考えに基づき、「知見」「特徴的変動」「Data Jacket」を連結したものをデータ利活用知識と定義する。「あ

連絡先: 高間康史, 首都大学東京大学院システムデザイン研究科, 〒 191-0065 東京都日野市旭が丘 6-6, ytakama@tmu.ac.jp

る知見がある動向情報の特徴的変動から発見された」ことを表す述語 (discover) 及び、「ある動向情報の特徴的変動がある動向情報から獲得された」ことを表す述語 (get) を定義し、モデル構築に利用する。この時、Data Jacket を作成した動向情報から特徴的変動を獲得した際に入力した「クエリ」をプロパティとして定義し、「特徴的変動」と「Data Jacket」を連結する際に付与する。この二つのモデルを図式化したものを図 1 に示す。Data Jacket 検索システムは、これらのモデルに基づき他者の知見などにアクセス可能とする。

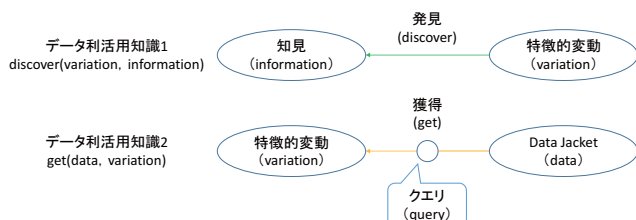


図 1: データ活用知識モデルの図式化

4. 評価実験

4.1 概要

工学系大学生・大学院生 10 名の実験協力者に、コンテキスト検索エンジンを用いて、あるアイテムと関係があると思われるアイテムを 5~15 分で可能な限り多く見つけてもらうタスクを行ってもらった。知見・特徴的変動・Data Jacket それぞれのデータベースから発見された Data Jacket の有用性を評価する 3 問のタスクを用意し、Data Jacket 検索システムなしのコンテキスト検索エンジンを用いて各タスクについて回答してもらった後、Data Jacket 検索システムありのコンテキスト検索エンジンを用いて同一タスクについてさらに回答してもらう。タスク内容を 2 パターン用意し、それぞれ実験協力者 5 人に回答してもらった。実験終了後、回答者とは別の実験協力者 2 人に、それぞれ 5 人分の回答の妥当性を 0~2 点で評価してもらい、1 点以上の場合は正解とし、0 点の場合は不正解とした。

それぞれのタスクについて見つけたアイテムの正解数を図 2, 図 3, 図 4 に示す。タスク 1 は、「指定された期間に急上昇したアイテムの一つ探し、そのアイテムと関係があると思われるアイテムを探す」というタスクであり自由度が高かったため、実験協力者 B, F, I, J は Data Jacket 検索システムなしでも比較的多数の正解アイテムを発見できている。しかし、10 名中 9 名の実験協力者では、Data Jacket 検索システムを

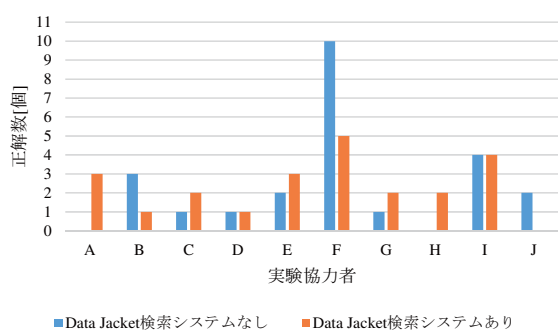


図 2: タスク 1 で見つけたアイテムの正解数

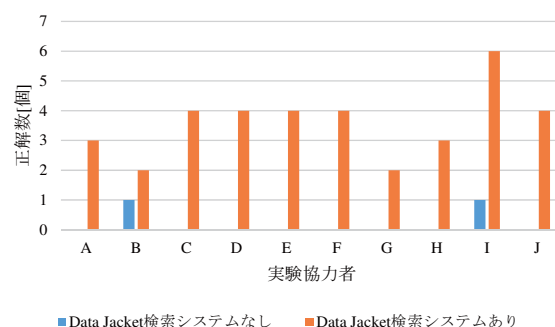


図 3: タスク 2 で見つけたアイテムの正解数

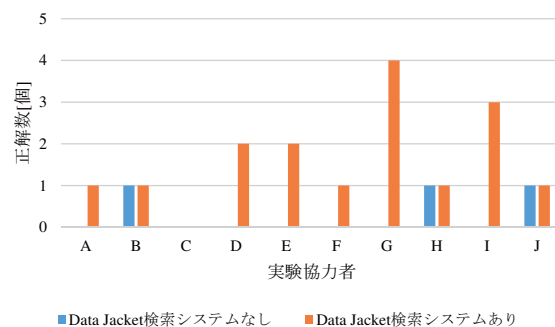


図 4: タスク 3 で見つけたアイテムの正解数

利用することで、さらにアイテムを発見している。特に 5 名の実験協力者においては、Data Jacket 検索システムを利用した場合の方が、より多くの正解アイテムを発見できている。また、他の 2 つのタスクは、「指定されたアイテムと関係があると思われるアイテムを探す」というタスクであったが、Data Jacket 検索システムなしの場合、関係のあるアイテムを見つけるのに苦労していたことがわかる。実験協力者 C のタスク 3 を除き、Data Jacket 検索システムを利用することで正解アイテムを発見できている。以上から、Data Jacket 検索システムを利用して過去の知見を頼りにすることで、関係があるアイテムの発見が容易になると考える。

5. おわりに

本稿では Data Jacket を導入し、ローカルデータと公開データを統合利用可能とするコンテキスト検索エンジンを提案した。評価実験の結果、Data Jacket 検索システムを利用することで、Data Jacket や過去に発見された知見を活用できることを示した。今後は、Data Jacket や知見の増加に対応可能な検索手法について検討する予定である。

参考文献

- [高間 15] 高間, 加藤, 桑折, 石川: 動向に関する問いを対象とした検索エンジンの提案, 人工知能学会論文誌, Vol.30, No.1, pp.138-147, 2015.
- [大澤 14] 大澤: データジャケット-創造的コミュニケーションのあるデータ市場のために-, 人工知能, Vol.29, No.6, pp.622-627, 2014.
- [早矢仕 16] 早矢仕, 大澤: Data Jacket Store: データ活用知識構造化と検索システム, 人工知能学会論文誌, Vol.31, No.5, pp.A-G15.1-9, 2016.