様式Ｆ－１９

科学研究費助成事業（学術研究助成基金助成金）研究成果報告書

平成 24 年 5 月 24 日現在

機関番号：12102
研究種目：若手研究(B)
研究期間：2011～2012
課題番号：23700102
研究課題名（和文）　オンライン分析による XML ストリームからの知識発見

研究課題名（英文）　Knowledge discovery over XML streams by online analysis

研究代表者
天笠　俊之（AMAGASA TOSHIYUKI）
筑波大学・システム情報系・准教授
研究者番号：70314531

研究成果の概要（和文）：
XML（Extensible Markup Language）は，標準データフォーマットとして広く利用されており，さらにセンサネットワークや M2M などの登場により，継続的に送信されるストリーム形式の XML データ（XML ストリーム）が増加しつつある．一方，膨大な XML ストリームに対して，単なる検索処理ではなく，より高度な分析処理を行ないたいという要求が存在する．このため本研究では，オンライン分析による XML ストリームからの知識発見手法を研究，開発した．

研究成果の概要（英文）：
XML (Extensible Markup Language) has widely been used as a standardized data format in wide spectrum of applications, and XML stream, where data records represented using XML are continuously transmitted as data streams, has been gaining its popularity due to the mergence and proliferation of sensor networks and M2M (Machine-to-Machine) applications. Meanwhile, the demand to perform analytical processing over XML streams rather than simple query retrieval is increasing. To this problem, in this research, we have studied the schemes for online analytical processing over XML streams.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
交付決定額	3,200,000	960,000	4,160,000

研究分野：データベース，データ工学
科研費の分科・細目：情報学・メディア情報学・データベース
キーワード： XML，XML ストリーム，分析処理

１．研究開始当初の背景

(1) XML (Extensible Markup Language) は，標準データフォーマットとして広く利用されている．その用途は Web 文書，ビジネス文書，科学データ，ビジネスデータ，ネットワークプロトコルなど多岐に渡り，現在の情報システムにおいて欠かすことができない重要な位置を占めている．その結果，膨大な情報資源が XML 形式で蓄積されつつあり，その量は加速度的に増加している．このため膨大な XML データからの情報抽出と知識発見は，情報資源の有効利用の観点から極めて

重要である．

(2) XMLデータから所望の情報を抽出するには， XQuery や XPath といった問合せ言語を用いるのが標準的な方法である．しかし，XML データの容量が膨大な場合や，高頻度かつ連続的に XML データが到着するような場合，従来型のアプローチでは不十分である．

２．研究の目的

上で述べた背景を受け，本研究ではオンライン分析による XML ストリームからの知識発見に関する研究開発を行うことを目的とする．具体的な研究の方法は以下の通り．

３．研究の方法

オンライン分析による XML ストリームからの知識発見のため，具体的には以下の三つの項目について研究開発を行なった：【研究項目 1:効率的な XML ストリーム処理手法】,【研究項目 2:XML データに対するオンライン分析手法】,【研究項目 3:XML データからの知識発見のための対話的分析手法】.

４．研究成果

(1) 【研究項目 1：効率的な XML ストリーム処理手法】では，XML ストリーム処理における構文解析処理のオーバーヘッドに着目した検索処理の効率化手法を開発した．具体的には，与えられた問合せに対して関係しない（結果に影響を与えない）部分 XML データについては，構文解析をスキップすることで処理効率の向上を図った．

図１は提案するシステムの概要図である．XPath 式で与えられる問合せ要求を事前に分析することによって，問合せの結果となり得ない要素が判断できる．そのような要素の場合には，問合せ処理器から XML パーサにスキップ処理の指示を出すことで処理を効率化できる．

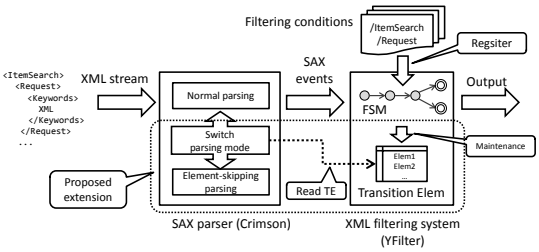

図 1：XML ストリーム処理の効率化.

(2) 【研究項目 2:XML データに対するオンライン分析手法】では，XML データに対して OLAP 分析を行うための手法を研究開発した．具体的には，XML 形式で表現されることの多い Linked Open Data（LOD）を対象として，LOD データに対して OLAP 分析を行うための手法を開発した．

本手法では，LOD データに対し OLAP 分析を行うためのスタースキーマをデータから生成する点に特徴がある．具体的には，LOD データのグラフ構造に着目した分析対象レコードの同定と，属性の抽出を行う．さらに，OLAP 分析のための概念階層を，1) LOD データ内のデータ型に由来するもの，2) LOD データ内に内在する階層構造，3) 外部データへの参照の三つに場合分けし，それぞれについて概念階層の取得方法を議論した．実験による評価により，実際の LOD データに対して OLAP 分析を行うことができることを示した．

(3) 【研究項目 3：XML データからの知識発見のための対話的分析手法】では，XML データに対して対話的な分析処理を花王にする手法を検討した．具体的には，対話的分析処理手法の一つであるファセット検索を対象として，XML データに対するファセットの抽出と，検索インタフェース構築のためのフレームワークを検討した．

まず，XML データに対して検索対象とな

る要素を抽出する．このため，XML データのスキーマ情報あるいは XML インスタンスから抽出した DataGuide を利用する．このようにして得られた XML データに対して，ファセットを定義する（図 2）．利用者の検索要求は，ファセットに対する種々の操作によって実現されるが，これを一般化するために利用者の対話操作に対応する演算を定義した．

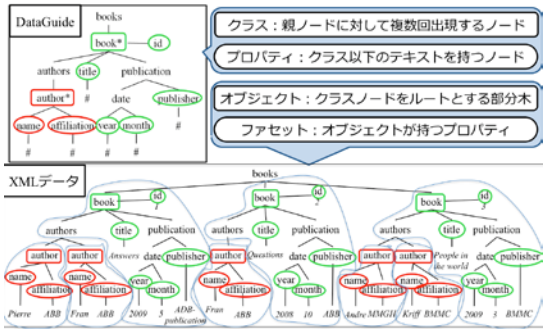

図 2：XML データに対するファセット．

提案したフレームワークはプロトタイプシステムとして実装し（図 3），その有効性はプロトタイプシステムを利用した被験者実験により示した．

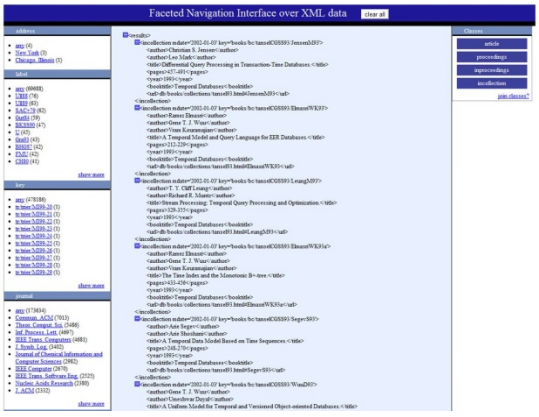

図 3：XML データのファセット検索．

５．主な発表論文等

（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 2 件）

(1) 井上寛之，天笠俊之，北川博之，LOD の OLAP 分析を可能にする ETL フレームワークの提案，日本データベース学会論文誌，Vol. 12, No. 1, 2013.（査読あり，印刷中）

(2) Takahiro Komamizu, Toshiyuki Amagasa, Hiroyuki Kitagawa, Faceted Navigation Framework for XML Data, International Journal of Web Information Systems (IJWIS), Vol. 8, No. 4, pp. 348-370, 2012.（査読あり）

〔学会発表〕（計 5 件）

(1) Hiroyuki Inoue, Toshiyuki Amagasa, and Hiroyuki Kitagawa, An ETL Framework for Online Analytical Processing of Linked Open Data (short), The 14th International Conference on Web-Age Information Management (WAIM 2013), Beidaihe, China, June 14-16, 2013.

(2) 井上寛之，天笠俊之，北川博之，LOD の OLAP 分析を可能にする ETL フレームワークの提案，第 5 回データ工学と情報マネジメントに関するフォーラム（DEIM フォーラム 2013），B2-3，福島県郡山市，2013 年 3 月 3 日．

(3) 井上寛之，天笠俊之，北川博之，“OLAP を利用した Linked Data の分析処理”，情報処理学会 第 74 回全国大会講演論文集，2012(1)，pp. 589-591，愛知県名古屋市，2012 年 3 月 6 日．

(4) 清野真奈，天笠俊之，北川博之，XML ストリームに対する省電力を考慮した問合せ処理，第 4 回データ工学と情報マネジメントに関するフォーラム（DEIM フォー

ラム 2012), E11-4, 兵庫県神戸市, 2012 年 3 月 5 日.

(5) Takahiro Komamizu, Toshiyuki Amagasa, Hiroyuki Kitagawa, A Framework of Faceted Navigation for XML Data, The 13th International Conference on Information Integration and Web-based Applications & Services (iiWAS2011), pp. 28-35, Ho Chi Minh City, Vietnam, 5-7 December, 2011.

６．研究組織

(1)研究代表者

天笠　俊之（AMAGASA TOSHIYUKI）

筑波大学・システム情報系・准教授

研究者番号：70314531