様式Ｃ－１９

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成２４年５月３１日現在

機関番号：12102
研究種目：基盤研究（Ｂ）
研究期間：2008～2011
課題番号：20300001
研究課題名（和文）　記号計算の理論を駆使したウェブソフトウェアのモデル化と検証

研究課題名（英文）　Modeling and verification of web software based on theories of symbolic computation

研究代表者
井田　哲雄　（IDA TETSUO）
筑波大学・システム情報系・教授
研究者番号：70100047

研究成果の概要（和文）：
ウェブソフトウェア検証の事例研究として，WebEos の核となる部分の形式化と検証を行った．幾何と代数の基本的な部分に Mathematica の計算結果を援用することで，効率的な検証が可能となった．文字列解析による検証において，正規表現マッチングの正確な解析を可能とした．また，データベースとの連携の解析を導入し，蓄積型 XSS 脆弱性検査を実現した．ポジションオートマトンを利用した正規表現の貪欲マッチングアルゴリズムの設計と実装を行った．

研究成果の概要（英文）：
As a case study of Web software verification, we have verified the core of WebEos. The effective verification was conducted by utilizing some results of the computation conducted on Mathematica. With respect to the verification based on string analysis, we developed the method to precisely analyze regular expression matching. By introducing the analysis of communication to database, we enabled the detection of stored XSS. We designed and implemented an algorithm of greedy regular expression matching based on position automata.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
2008 年度	5,500,000	1,650,000	7,150,000
2009 年度	3,600,000	1,080,000	4.680,000
2010 年度	3,200,000	960,000	4,160,000
2011 年度	2,400,000	720,000	3,120,000
総　計	14,700,000	4,410,000	19,110,000

研究分野：情報科学
科研費の分科・細目：情報学、情報学基礎
キーワード：ソフトウェア検証，記号計算，ウェブ

１．研究開始当初の背景

インターネットの急速な普及と発展に伴って，ソフトウェアが XML ドキュメントに典型的に見られるようなプログラムとデータが混在した構造物（本研究ではこれらをウェブソフトウェアと呼ぶ）に変容しつつある．ウェブソフトウェアには従来にない高い信頼性が要求される．単に誤りなく動作するだけでなく，頑健である，悪用・誤用されないといった性質がウェブソフトウェアに求められている．このような「正しさ」を保証するには従来の開発手法やデバッグ手法だけでは不十分であり，産業界や多くの大学，研究機関で新たな方法論や方法論を支援する言語やツールが作られてきている．マイクロソフト社の Static Driver Verifier，グーグル

社の GWT(Google Web Toolkit)，ミュンヘン工科大学のニプコフらの Isabelle/HOL，INRIA の Objective Caml と Coq，カーネギメロン大学のクラークらの SMV などが代表的である．

２．研究の目的

記号計算の理論，オートマトン理論をはじめとする，計算やソフトウェアを形式化する様々な理論を駆使して，上記問題の解決に取り組む．我々は，定理証明支援系やモデル検査系を用いた方法論の高度化と広範な展開が重要であると考え，これまで証明支援系を用いたソフトウェアや幾何オブジェクト設計の正当性証明の研究を推進してきた．それと同時に，記号計算を支援するウェブサービスのシステムやサーバーサイドの記号計算アルゴリズムをクライアント（ブラウザー）から対話的にアクセスするウェブソフトウェアを構築してきた．これらの先行研究で我々が得てきた知見や経験及び記号計算の研究コミュニティーが蓄積してきた多くの知見をウェブソフトウェアの「正しさ」の検証に活用して，新たな理論，方法論，および方法論を支援するツール群を構築する．

３．研究の方法

複雑なウェブソフトウェアを研究の対象にするので，ウェブソフトウェアの研究を次の三つに分けて考究する．

- システムとしてのモデル化と検証（担当　南出　井田）
- ウェブと交信し，サーバー側あるいはクライアント側で稼働するプログラムのモデル化と検証　（担当　井田　南出）
- XML ドキュメントの検証　（担当　マリン，鈴木）

ウェブソフトウェアを解析し，3 つのモデルに分けて研究を進める．モデル化は研究のすべての局面で必要となる研究手法であり，モデル化のサイクルを繰り返しモデルを洗練させる．また，問題に適したモデル化や，一つの問題に対して複数のモデル化を考える必要がある．最終的には，モデルを既存の論理体系，計算体系におとして，必要とされる性質の検証を行う．検証の対象や性質に応じて，いくつかの定理証明支援系やモデル検査系を用いる．我々は，Isabelle/HOL，Theorema，SMV を研究のみならず，教育にも用いており，これらのシステムを活用する．さらに，証明ライブラリーの豊富な Coq や数学の定理証明に豊富な蓄積のある Mizar も適宜参考にする．

４．研究成果

ウェブソフトウェアのモデル化と検証に様々な角度から取り組んだが，研究活動は次の６点に集約することができる．以下に，研究成果の概要を述べる．具体的な成果は，公表した論文を参照されたい．

（１）プログラムの性質を表す論理式を動的に生成し，生成された論理式の正しさを証明する研究を推進した．特に，シリンダー代数分解やグレブナ基底計算法の応用研究を行った．グレブナ基底計算の単項順序を工夫することにより，幾何学的な証明問題が効率よく行われることが判明したので，多くの実例についてこれを実証するとともに，ウェブプログラムに対して，応用の方策を検討した．

（２）WebEos(Web E-Origami System の略)と呼ぶウェブを介して稼働するソフトウェアを開発してきたが，本研究においても，このソフトウェアの改良とその核となる部分の形式化を進めた．核となる部分は，藤田の基本操作と呼ばれるセットをプログラム化したものであり，この形式化と様々な性質の検証が非常に重要であった．これを，自動定理証明支援系 Isabelle/HOL を用いて行った．検証には，記号論理的な知見と代数的な知見を組み合わせることが不可欠であるが，既存の自動定理証明支援系は，前者では満足すべき能力を有するものの，後者は比較的弱く，研究の推進には困難が伴った．得られた検証スクリプトは，様々なバージョンを含めると，数千行にもなるが，幾何と代数の基本的なところは独自の証明スクリプトをベースにするのではなく，記号代数系 Mathematica の計算結果を援用した．これらの研究成果により，既に稼働しているソフトウェアの改善及び機能の向上が得られ，ウェブソフトウェアの質のより一層の向上が得られた．

（３）ウェブソフトウェアの開発で用いられる代表的スクリプト言語である PHP および Ruby の意味論の研究を行った．PHP については，その特徴となる参照代入に注目し，その意味論をグラフ書換として定式化し，現行の実装の問題点を明らかにした．また，Ruby については，そのオブジェクト言語機能を表現する操作的意味論を構築した．この操作的意味論に基づき，プログラムの制御フローを識別するコントロールフロー解析手法を開発した．

（４）文字列解析によるウェブプログラム検証の実用性を高める研究に取り組んだ．データベースとの連携の解析を導入することにより，文字列解析による蓄積型クロスサイトスクリプティング脆弱性検査を実現した．また，クロスサイトスクリプティング脆弱生の検査と反例生成の仕組みを，トランスデューサを用い

て再構成し柔軟な検査を可能にした.

（５）スクリプト言語のプログラムにおいて重要な役割を果たす正則表現マッチングの研究を行った.多くのスクリプト言語が採用している欲張り戦略による実装の意味をモナドの概念を用いて定式化し,マッチングに精密に対応する出力付きオートマトンを構成する方法を与えた.この出力付きオートマトンを用いることで,文字列解析の精度を高めることができる.さらに，構築したモナドを用いた意味論を，モナドモルフィズム，モナドトランスフォーマを用いて再構成し,より洗練された意味論を構築した.

（６）XML 文書の効率的な文書変換に関する研究を推進した． XML 文書と変換規則との効率の良いマッチングアルゴリズムを,より広いクラスの変換規則に適用することを目指した.ヘッジではなくシーケンスに関する正規表現から検討を開始し，POSIX 準拠のマッチング及び欲張りマッチングの両方について,効率的なマッチングアルゴリズムの設計した.正規表現の貪欲マッチングのためのアルゴリズムの設計においては,ポジションオートマトンとよばれるオートマトンを利用した.アルゴリズムが望ましいマッチング結果を返すことを検証するために,このオートマトンの各遷移に正規表現中のキャプチャの位置を表すタグを付加した.マッチングの過程で通った遷移がもつタグの列を作り,それらの間の順序を比較することで,貪欲マッチングで得られるべきキャプチャの結果を正しく得られることを示した.それをもとにアルゴリズムの設計と C++による実装を行った.その結果いくつかの場合でFrischとCardelliによって提案された効率のよいマッチングアルゴリズムの速度を大きく上回る結果が得られた.

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１３件）

① Yuto Sakuma, Yasuhiko Minamide, Andrei Voronkov, Translating Regular Expression Matching into Transducers, Journal of Applied Logic, 査読有, 10, 2012, 32-51

② Taro SUZUKI, Junya TERAZONO, Takafumi HAYASHI, Design and Implementation of New Generation Data Format for Lunar and Planetary Exploration, JSASS on-line journal ″Aerospace Technology″ (to appear), 査読有, 2012

③ Ida. T, Kasem. A, Ghourabi. F, Takahashi. H, Morley's theorem revisited: Origami construction and automated proof, Journal of Symbolic Computation, vol. 46, pp. 162-170, 2011, 査読有

④ Kaliszyk. C, and Ida. T, Proof Assistant Decision Procedures for Formalizing Origami, Lecture Notes in Computer Science (proceedings of the Conference on Intelligent Computer Mathematics (CICM'11)), 査読有, 6824, 2011, 45-57.

⑤ Ghourabi. F, Ida. T, and Kasem. A, Proof Documents for Automated Origami Theorem Proving, Lecture Notes in Computer Science (post-proceeding of the 8th International Workshop on Automated Deduction in Geometry (ADG 2010)), 査読有, 6877, 2011,78-97.

⑥ Kasem. A, Ghourabi. F, Ida. T, Origami Axioms and Circle Extension, Proceedings of the 26th Symposium on Applied Computing (SAC 2011), pp. 1106-1111, 2011,査読有

⑦ Tetsuo Ida and Hidekazu Takahashi. Origami Fold as Algebraic Graph Rewriting. Journal of Symbolic Computation, 45(4):393 - 413, 2010. 査読有

⑧ 松本宗太郎, 南出靖彦, Ruby プログラムの制御フロー解析とその健全性の証明, 3 巻, pp. 9-25,2010.査読有

⑨ Tozawa, M. Tatsubori, T. Onodera, Y. Minamide, Copy-on-Write in the PHP Langauge, Proc. of the 36th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, pp.200-212, 2009, 査読有

⑩ Tetsuo Ida. Symbolic and Algebraic Methods in Computational Origami: Invited Talk. In Proceedings of the 2009 International Symposium on Symbolic and Algebraic Computation (ISSAC 2009), pages 3-4. ACM, 2009. 査読有

⑪ Tetsuo Ida, Hidekaza Takahashi, Origami Fold as Algebraic Graph Rewriting, Proc. of 24th annual ACM Symposium on Applied Computing, pp. 1132-1138, 2009, 査読有

⑫ T. Nishiyama, Y. Minamide, A Translation from the HTML DTD into a Regular Hedge Grammar, Proc. of 13th

International Conference on Implementation and Application of Automata, pp. 122-131, 2008, 査読有

⑬ Taro Suzuki, Staoshi Okui, Product Drivatives of Regular Expressions, IPSJ Online Transactions, 1 巻, pp.53-65,2008, 査読有

〔学会発表〕(計５件)

① 木村 将人, 南出 靖彦, 文字列解析によるクロスサイトスクリプティング脆弱性検査の改良, プログラミングおよびプログラミング言語ワークショップ(ポスター), 2012 年 3 月 8 日, 和歌山県 南紀白浜

② Kasem and T. Ida, Extended Web Services for Computational Origami. The 3rd International Workshop on Symbolic Computation in Software Science (SCSS2010), RISC-Linz Report Series No. 10-10, pages 144-154. Hagenberg, Austria, July 30, 2010.

③ 松本宗太郎, 南出靖彦, Ruby のコア言語の操作的意味論, 日本ソフトウェア科学会第 26 回大会, 2009 年 9 月 16 日, 島根大学

④ F. Ghourabi, T. Ida, H. Takahashi, and A. Kasem. Reasoning Tool for Mathematical Origami Construction. The International Symposium on Symbolic and Algebraic Computation (ISSAC 2009), Seoul, Korea, July 30, 2009.

⑤ Kasem and T. Ida. Experiences with Web Environment Origamium. 日本ソフトウェア科学会第２５回大会, 筑波大学(東京キャンパス), 2008 年 9 月 11 日

６．研究組織

(1)研究代表者

井田　哲雄　(IDA TETSUO)
筑波大学・システム情報系・教授
研究者番号：70100047

(2)研究分担者

南出　靖彦　(MINAMIDE YASUHIKO)
筑波大学・システム情報系・准教授
研究者番号：50252531

Marin Mircea　(MARIN MIRCEA)
筑波大学・システム情報系・講師
研究者番号：60396603

鈴木 大郎　(SUSUKI TARO)
会津大学・コンピュータ理工学部・准教授
研究者番号：90272179