

ブログ上の広告活動を対象としたアフィリエイト分析支援システム

石井 聡[†] 福原 知宏^{††} 増田 英孝[†] 中川 裕志^{†††}

[†] 東京電機大学未来科学部 〒101-8457 東京都千代田区神田錦町 2-2

^{††} 東京大学人工物工学研究センター 〒277-8568 千葉県柏市柏の葉 5-1-5

^{†††} 東京大学情報基盤センター 〒113-0033 東京都文京区本郷 7-5-1

E-mail: †ishii@csl.im.dendai.ac.jp

あらまし 近年、ブログやホームページ上に企業の広告を掲載する成功報酬型広告（アフィリエイトプログラム）が盛んに行われている。アフィリエイトプログラムにより、個人が手軽に Web 上で紹介したい商品の情報を掲載できるようになった。一方で、広告収入を目的としたスパム（アフィリエイトスパム）が問題となっている。先行研究では、アフィリエイトスパムに関する研究は行われているが、実際に個人の Web サイトでどのような広告活動が行われているかは明らかではない。本研究では、アフィリエイト識別番号（アフィリエイト ID）に着目したブログ上での広告活動分析アプローチ（アフィリエイト ID 特定分析）を提案し、アフィリエイト ID 単位での記事数、掲載商品数、商品ジャンル、ブログ数を調査するアフィリエイト分析支援システムの開発を行った。

キーワード ブログ, アフィリエイト

A System for Analyzing Affiliate Information on Blogs

Soichi ISHII[†], Tomohiro FUKUHARA^{††}, Hidetaka MASUDA[†], and Hiroshi NAKAGAWA^{†††}

[†] School of Science and Technology for Future Life, Tokyo Denki University Nichikicho 2-2, Chiyoda-ku, Tokyo, 101-8457 Japan

^{††} Research into Artifacts Center for Engineering, The University of Tokyo Kashiwanoha 5-1-5, Kashiwa-shi, Chiba, 277-8568 Japan

^{†††} Information Technology Center, The University of Tokyo Hongo 7-5-1, Bunkyo-ku, Tokyo, 113-0033 Japan

E-mail: †ishii@csl.im.dendai.ac.jp

Abstract In this paper, we report analysis results of affiliate marketing on the Web. Recently, affiliate programs such as Rakuten affiliate and Amazon associates program became popular on various web sites including personal blog sites. By using affiliate programs, bloggers and webmasters can easily add advertisements on their blog article and web pages, and people who access to these pages can find products related to contents those web pages. Our aim to find the state of affiliate marketing on the web; how many people incorporate affiliate programs into their web sites, what kind of products they sell and how much do they sell those products.

Key words Blog, Affiliate Program

1. はじめに

近年、ブログやホームページ上に企業の広告を掲載する成功報酬型広告（アフィリエイトプログラム）が盛んに行われている。個人が手軽に Web 上で紹介したい商品の情報を掲載できるようになり、今までのようなメーカーや販売店からの商品情報だけでなく、利用者の商品レビューなどを簡単に得られるようになった。一方で、アフィリエイトプログラムにより、個人の Web サイトまでも広告媒体として広く認知され、広告収入を

目的としたスパム（アフィリエイトスパム）が問題となっている。

我々はアフィリエイトスパムのフィルタリングを目的として、アフィリエイト識別番号（アフィリエイト ID）に着目した分析アプローチを提案する。先行研究 [1] [2] での Web ページや Web サイトを基準とした分析手法では分析できない複数のブログサービスにまたがったアフィリエイトの分析や、掲載商品を調査する Web システムを開発する。これによって、今まで明らかでなかったアフィリエイト空間の分析を行う。

本分析では、アフィリエイト ID 単位での分析となるので、

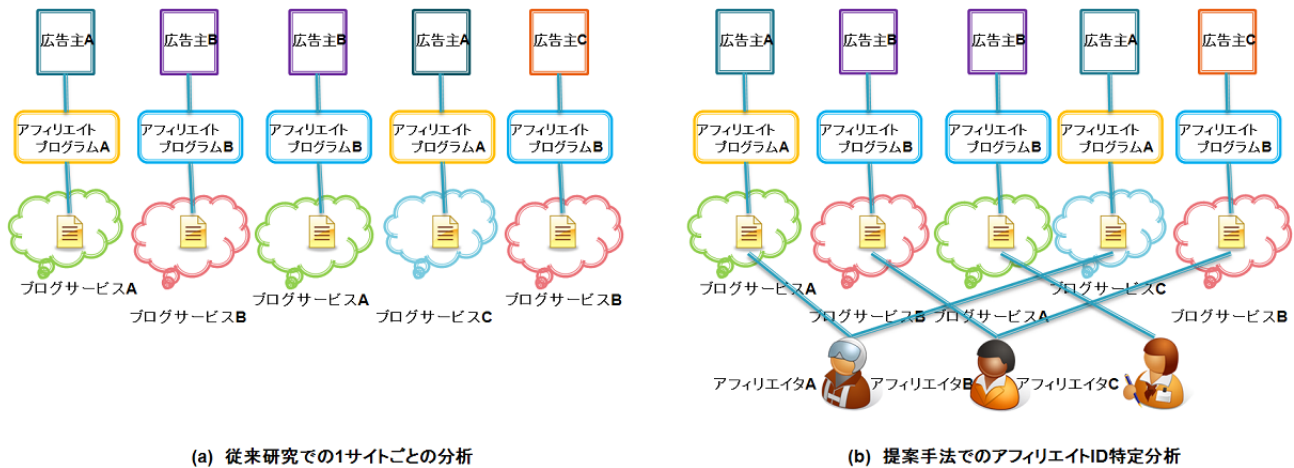


図 1 先行研究の分析手法とアフィリエイト ID 特定分析手法

Web サイト単位では特定することが難しかった複数のブログを管理するアフィリエイトの商品傾向や、商品ジャンル、価格帯も調べることが可能となる。

本論文の構成は次の通りである．2. で先行研究の分析手法と我々の提案する分析アプローチを示す．3. でシステムの説明を行い．4. で実際にシステムを利用した分析事例を紹介する．5. で考察を述べ、6. で本論文のまとめと今後の課題について述べる．

2. 先行研究の分析手法と本研究の分析手法

2.1 先行研究

アフィリエイトスパムのフィルタリングに関して、長谷 [1] らの研究では 2 つの分析が行われている．(1) ブログ内に含まれるアフィリエイトリンク数の集計、(2) スパムブログとアフィリエイトプログラムとの関係性の調査である．長谷らの分析結果では、1 つのブログ内に含まれるアフィリエイトリンク数が多いほど、スパムブログ (Splog) である傾向が高く、アフィリエイトブログの半数がスパムブログであると報告している．また、スパム割合が 90% 以上のアフィリエイトプログラム 5 社の存在を報告している．

Wang [2] らの研究では、スパムサイトから広告主に至るまでのトラフィック分析を行っている．アフィリエイトスパムのネットワークが、

- 第 1 層 Doorway
- 第 2 層 Redirection Domain
- 第 3 層 Aggregators(プロバイダ)
- 第 4 層 Syndicator(ASP(Affiliate Service Provider))
- 第 5 層 Advertiser(広告主)

の 5 階層に渡っており、Doorway と Advertiser が多く存在し、Redirection Domain と Syndicator が少なく、Aggregators がさらに少数であることを示した．それが 2 つの漏斗を上下逆さまにして繋げたような形となることから Spam Double-Funnel モデルと名付けている．この各階層において頻出するドメインの統計分析を行っている．

斉藤 [3] らの研究では、ブログ中の商品紹介記事を収集し、ア

フィリエイトの有無によって、言語表現に異なる特徴が現れるか検討を行っている．この結果、アフィリエイトのある記事には、書き手の意見や印象を強調する表現が多く現れていることを報告している．

2.2 本研究におけるアプローチ

図 1 に我々が提案する分析手法を示す．本研究ではアフィリエイト ID 特定分析に基づき、アフィリエイト空間の分析を行う．アフィリエイト ID 特定分析とは、ASP(Affiliate Service Provider)^(注1)から登録ユーザ各々に付与される独自の ID(アフィリエイト ID) をブログ記事等から抽出し、アフィリエイト ID を単位として分析する分析手法である．

先行研究では、1 つのウェブページやウェブサイトを対象とし、そのウェブページのドメインや利用するアフィリエイトプログラム、掲載する広告主といった縦の関係のトラフィック分析 (ブログサイト毎の独立した分析) を行っていた (図 1(a))．これに対し、我々の提案するアフィリエイト ID 特定分析では、先行研究のトラフィック分析に加え、アフィリエイト記事に含まれるアフィリエイト ID を用いて情報発信者 (アフィリエイト) を特定する (図 1(b))．これにより、どのアフィリエイトがどのブログで広告活動を行っているのかという情報を得ることができ、アフィリエイトスパムに見られる同一のアフィリエイト ID を用いた複数のブログサイトにまたがるスパムを特定できる．この結果、従来の分析方法では得られなかった横の関係の分析 (アフィリエイト ID からブログサービスが異なるブログサイトであっても、同じユーザのブログとして分析) も可能となる．

我々は、アフィリエイト ID 特定分析に基づき、アフィリエイトスパムの分析を行った [4]．その結果、アフィリエイトが管理するブログサイト数においてスパム、非スパムで顕著な差を示し、スパムでは複数のブログサイトを管理するアフィリエイトが 67.9%、非スパムでは 4.5%であることを報告した．

本論文で提案する分析アプローチでは、1 つのアフィリエイト ID を 1 ユーザとみなす．アフィリエイトでは通常、継続的

(注1): アフィリエイトプログラム提供会社、広告主と広告掲載主との間にいる中間業者である．

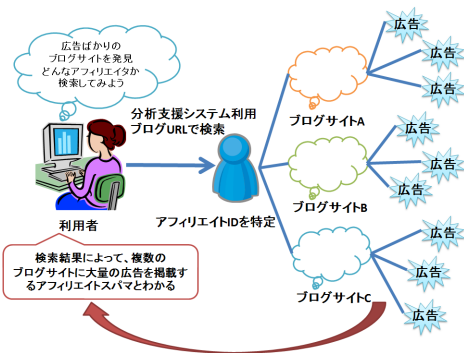


図 2 アフィリエイト分析支援システム利用イメージ図

に実績をあげていると紹介料率^(注2)が上がるなどの優遇措置がある[5]。このことからユーザは1つのアフィリエイトサービスに付き1つのアフィリエイトIDを取得している可能性が高いと考えられる。例外として、Amazonでは、1つのIDにつき複数のトラッキングIDを利用することができ、複数のアフィリエイトIDの収益を1つのアフィリエイトIDにまとめることが可能である。

2.3 アフィリエイトIDに着目したアフィリエイト検索

アフィリエイト空間分析のために、アフィリエイトID特定分析を用いたアフィリエイト検索を提案する。本システムに求められる要件を以下に定義する。

(1) 一般アフィリエイトとスパムアフィリエイトの活動比較

一般アフィリエイトとスパムアフィリエイトで広告掲載活動、掲載商品にどのような差があるのかの比較を行う。比較のために、商品の価格帯、商品カテゴリ、商品ジャンルを解析する機能が必要である。

(2) 人気アフィリエイトサイトに共通する特徴の分析

人気のあるアフィリエイトサイトに共通する特徴があるのか分析を行う。人気のあるブログサイトのURLを登録すると、記事を全て取得し、解析する機能が必要である。

(3) 複数ASP利用者の使い分け

図3にイメージ図を示す。複数のアフィリエイトプログラムを利用するアフィリエイトはどのような使い分けを行っているのかの分析を行う(例: Amazonでは、本、音楽、DVDを主に紹介し、楽天では、食品、アパレルなど)。複数のブログサイトを管理するアフィリエイトのブログサイトに、他のアフィリエイトIDを見つけた場合は、その2つのIDを1人のアフィリエイトとして解析する機能とAmazon以外の商品分析を行う必要がある。

(4) アフィリエイトの時系列分析

スパムアフィリエイトIDはどれぐらいの期間利用されているのか、また、アフィリエイトスパムを1週間、1ヶ月、1年というスパンで観察することによって、どのような活動(掲載商品のジャンル、価格帯、掲載頻度など)を行っているのか分析を行う。このためには継続的なデータ取得と解析、また記事投稿間隔を調べる機能が必要である。

(注2): アフィリエイトから商品が購入された場合に、広告掲載主に販売額の何%を支払うかを定める値。

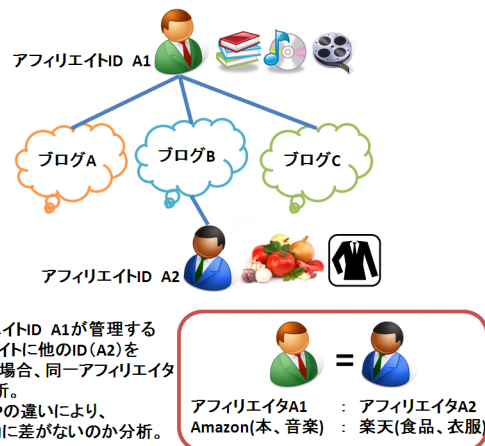
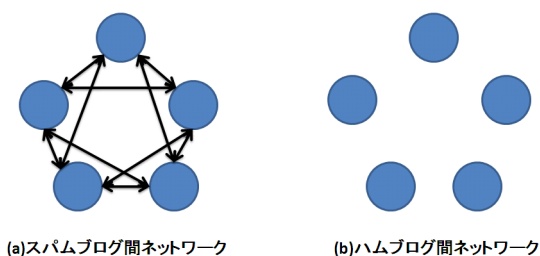


図 3 複数 ASP 利用者の使い分けイメージ図



アフィリエイトスパムはSEO対策の為に管理する複数のブログサイトで相互リンク

一般アフィリエイトは同じユーザがブログサイトを管理していると知られたくないので、リンクしない。

図 4 ブログサイト間のリンク解析イメージ図

(5) ブログ記事の本文分析

ブログ記事の本文抽出を行い、記事本文に広告に対するレビューがあるのか、それともコピー＆ペーストなのか内容分析を行う。また、本文抽出によって、広告が記事本文にあるのか、それ以外の部分(サイドバーなど)にあるのか、およその広告掲載位置を調べることが可能となる。

(6) 予約商品の分析

人気商品をいち早く紹介するアフィリエイトはどういったアフィリエイトなのか(人気、一般、スパム)分析を行う。記事更新日と商品発売日の差分を取ることで、発売前から商品を紹介しているのか調べることが可能である。

(7) ブログサイト間のリンク解析

図4にイメージ図を示す。図4(b)のハムとは非スパムのことである。Web上においてアフィリエイトブログがどのようなリンク構造をしているのか、リンク解析を用いて分析を行う。目視では、1つのアフィリエイトIDで約150のブログサイトを管理し、自分が管理するブログサイトに相互リンクを貼っていたアフィリエイトの存在を確認した。

本論文では、この中から(1)の一般アフィリエイトとスパムアフィリエイトの活動にどのような差があるのかを比較するための機能を実装した。

図2に検索イメージ図を示す。利用者はアフィリエイトID(URL、商品名でも可)をクエリとして、ブログサイト、掲載商品を検索することが可能である。この検索により、そのア

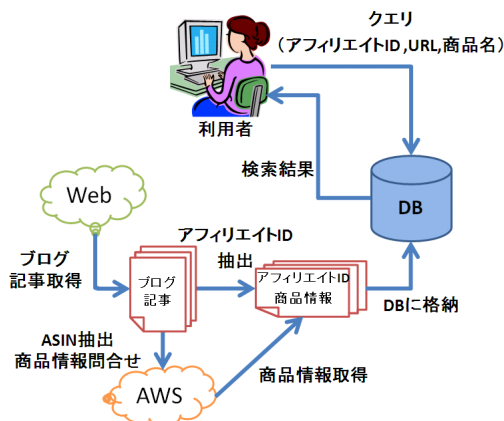


図5 システム構成

フィリエイト ID が存在するブログサイト、今まで扱ってきた掲載商品、掲載商品カテゴリ、ジャンルを得ることができる。これによって、アフィリエイトがいつどんな商品をどのように紹介していたかが分かる。例えば、アフィリエイトスパムが大量に記事を投稿している場合、その記事にあるアフィリエイトは全て同じ商品であるのか、または、ブログサイトによって掲載する商品を変えているのかなどがわかり、アフィリエイトスパムの活動内容の分析が可能である。

3. アフィリエイト分析支援システム

3.1 分析データ

分析データとして、ブログサービス運営会社8社^(注3)から取得したブログ記事を利用する。提案システムはブログサービスから配信されるRSSファイルを定期的に取得し、そこから新着記事のURLを抽出し、ブログ記事を収集する。クロールは1日おきに0時、12時、23時台の記事を取得している。また、取得した記事からHTML内にあるリンク、HTMLタグ情報などを用いて、ASP13社の^(注4)アフィリエイトプログラムの記事数を集計し、その中でも記事数の多かった上位3社として、Amazonアソシエイト(以下Amazon)、Google AdSense(以下Google)、楽天アフィリエイト(以下楽天)のアフィリエイトの選定を行った。今回の分析では、商品分析はAmazonだけとし、商品情報取得のためにAmazonを利用する記事からAmazon独自の商品識別番号ASIN(Amazon Standard Identification Number)を取得し、AWS(Amazon Web Services)に問い合わせ商品情報を取得する。

3.2 システム構成

図5にシステム構成を示す。1台の計算機を用いて、Ubuntu 9.04, MySQL 5.0.75, Apache 2.2.11 を利用している。提案システムはWebから毎日ブログ記事を収集し、そこからアフィリエイトIDとASINの抽出、問合せを行いデータベースに格

(注3): Ameba, fc2, JUGEM, livedoor, Seesaa, Yahoo!, ココログ, ウェブブログ

(注4): A8.net, AccessTrade, Amazon アソシエイト, DMM, Google AdSense, infocart, infostore, infostyle, LinkShare, Store-mix, TrafficGate, ValueCommerce, 楽天アフィリエイト

商品数	最高値	最安値	平均価格
204	80000円	281円	1134円

商品ジャンル	カテゴリー	件数
ヘルス&ビューティー	ヘルス&ビューティー-Refinements	35
ジャンル別	ジャンル別-Refinements	31
科学・テクノロジー	科学・テクノロジー-Refinements	10
旅行・観光	旅行・観光-Refinements	10
ホーム&キッチン	ホーム&キッチン-Refinements	15
おもちゃ&ゲーム	おもちゃ&ゲーム-Refinements	12
スポーツ&アウトドア	スポーツ&アウトドア-Refinements	11
エンターテインメント	エンターテインメント-Refinements	10
ファッション	ファッション-Refinements	10
ペット&ペット用品	ペット&ペット用品-Refinements	9

ブログURL	記事URL	商品ID	商品名	記事更新日	記事取得日	ASP
http://www.livedoor.biz	http://www.livedoor.biz/archives/049228.html	4062784164	神楽坂 コインロッカー・ペイジー (講談社文庫)	2009-12-19 22:43:10.0	2009-12-20 01:56:41.0	Amazon
http://www.livedoor.biz	http://www.livedoor.biz/archives/049228.html	8000JA7080	古銭入門 古銭の収集の手引 (1959年)	2009-12-19 22:43:10.0	2009-12-20 01:56:41.0	Amazon

図6 アフィリエイト ID 検索結果画面

納している。2010年1月時点で、アフィリエイト記事約10万件、商品件数約8万件のデータが蓄積されている。

4. アフィリエイト分析支援システムを用いた分析事例

アフィリエイト分析支援システムを実際に利用した結果を、システムの出力画面と共に説明する。図6にアフィリエイトIDをクエリとした時の出力画面を示す。出力画面はメニューバー、商品一覧ジャンルテーブル、記事一覧テーブルの4つのブロックに分けられる。メニューバー部分ではブログサイト一覧がドロップダウンメニューになっており、クエリのアフィリエイトIDが管理しているブログサイトを一覧で表示し、ブログサイトを基準とした分析結果を表示するページにリンクされている。価格帯一覧では、クエリのアフィリエイトIDが管理するブログサイト全ての商品数、最高値、最安値、平均価格を表示している。商品カテゴリー一覧では、アフィリエイトIDが扱っている商品ジャンル上位10件と商品カテゴリ(例:本、音楽、DVD)から商品ジャンル(例:ミステリー、J-POP、アクション)までをパスとして表示している。記事一覧では、ブログURL、記事URL、商品ID、商品名、記事更新日、記事取得日、ASPを表示しており、どの記事でどの商品が紹介されているのかわかる。また、列毎にソート機能が付いているので、記事更新日でソートすることによって、同時刻帯に複数のブログサイトに記事を投稿しているアフィリエイトの分析が可能となる。

7つのブログサイトを管理するスパムアフィリエイトIDをクエリとし、本システムで分析を行い、従来研究での分析手法と提案するアフィリエイトID特定分析での商品ジャンル分析の比較を行う。図7に従来研究でのブログサイト毎(7サイトの中から3サイト)の商品ジャンル上位10件の結果を、図8にアフィリエイトID特定分析でのアフィリエイトID毎の商品ジャンル上位10件の結果を示す。従来研究では、複数のブログサイトを管理するアフィリエイトのブログサイトであっても、1人のアフィリエイトによって管理されていることを特定することが難しく、ブログサイト単位で独立して分析を行っていた。図7からブログサイト単位で商品ジャンルの分析を行うと、ブログサイトごと商品ジャンルに統一性が見て取れる。提案手

アフィリエイトA ブログ1		
商品ジャンル	カウント数	パス
★ヘルス &ビューティー	17	/ヘルス &ビューティー /Refinements
★ジャンル別	8	/本 /ジャンル別 /雑誌
★出版社	4	/本 /ジャンル別 /コミック・アニメ・BL /出版社別 /講談社 /キスコミックス
★カテゴリ別	3	/ヘルス &ビューティー /カテゴリ別 /成人向け・アダルトグッズ
★ヘルスクエアグッズ	3	/ヘルス &ビューティー /カテゴリ別 /ヘルスクエア /ヘルスクエアグッズ /ガルマニウム
★日用品	2	/ヘルス &ビューティー /カテゴリ別 /日用品 /その他
★掃除グッズ	2	/ヘルス &ビューティー /カテゴリ別 /日用品 /掃除グッズ /その他
★ヘルスクエア	1	/ヘルス &ビューティー /カテゴリ別 /ヘルスクエア /その他

アフィリエイトA ブログ2		
商品ジャンル	カウント数	パス
★おもちゃ&ホビー	12	/おもちゃ&ホビー /Refinements
★ホビー	7	/本 /ジャンル別 /実用・スポーツ・ホビー /ホビー /切手・コイン
★ホビー	6	/おもちゃ&ホビー /カテゴリ別 /ホビー /切手
★文具・ステーショナリー	5	/おもちゃ&ホビー /カテゴリ別 /雑貨・インテリア /文具・ステーショナリー /その他
★A-Z	4	/おもちゃ&ホビー /By Characters & Brands /A-Z /SIG
★ジャンル別	3	/本 /ジャンル別 /実用・スポーツ・ホビー
★家電&カメラ	2	/家電&カメラ /カテゴリ別

アフィリエイトA ブログ3		
商品ジャンル	カウント数	パス
★科学・テクノロジー	16	/本 /ジャンル別 /科学・テクノロジー /科学・テクノロジー 全般
★生物・バイオテクノロジー	9	/本 /ジャンル別 /科学・テクノロジー /生物・バイオテクノロジー /動物学
★ペット	8	/本 /ジャンル別 /暮らし・健康・子育て /ペット /ペット 一般
★ジャンル別	2	/本 /ジャンル別 /エンターテインメント

図 7 従来研究：ブログサイトごとでの商品ジャンル分析

アフィリエイトA		
商品ジャンル	カウント数	パス
ヘルス &ビューティー	35	/ヘルス &ビューティー /Refinements
★ジャンル別	31	/本 /ジャンル別 /雑誌
★科学・テクノロジー	16	/本 /ジャンル別 /科学・テクノロジー /科学・テクノロジー 全般
★旅行ガイド	16	/本 /ジャンル別 /旅行ガイド /旅行ガイド 全般
★ホーム&キッチン	15	/ホーム&キッチン /By Brands
★おもちゃ&ホビー	12	/おもちゃ&ホビー /Refinements
★実用・スポーツ・ホビー	11	/本 /ジャンル別 /実用・スポーツ・ホビー /実用・スポーツ・ホビー 全般
★エンターテインメント	10	/本 /ジャンル別 /エンターテインメント /エンターテインメント 全般
★カテゴリ別	10	/ホーム&キッチン /カテゴリ別 /キッチン&テーブルウェア
★生物・バイオテクノロジー	9	/本 /ジャンル別 /科学・テクノロジー /生物・バイオテクノロジー /動物学

図 8 提案手法：アフィリエイト ID での商品ジャンル分析

法のアフィリエイト ID 特定分析では、アフィリエイト ID が管理するブログサイト（今回は 7 サイト）で扱っている商品をまとめて商品ジャンルの分析を行っている。図 8 のアフィリエイト ID 単位での商品分析を見ると、おもちゃ&ホビーから科学・テクノロジーまでと商品ジャンルが多種に渡っていることがわかる。また、図 7 それぞれの上位にあるヘルス&ビューティー、おもちゃ&ホビー、科学・テクノロジーが図 8 にもあることがわかる。これからこのアフィリエイトは複数のブログサイトを管理し、かつブログサイトごとに商品傾向を変えている可能性が考えられる。

5. 考 察

本分析手法を用いたアフィリエイト分析支援システムで行った分析により、ブログサイト毎に掲載商品ジャンルに統一性を持たせるアフィリエイトスパムの存在を確認した。目視で、7 つのブログサイトを確認すると、ブログサイト毎に 1 つのキーワードに沿って記事や広告が掲載されていた。予備実験では、スパムアフィリエイトの方が、商品ジャンル、価格帯の幅が広いという結果を確認している。今後、定量的な調査を行いアフィリエイトスパムに共通した広告掲載活動があるのか分析を行う必要がある。

6. お わ り に

本論文では、個人のアフィリエイトが Web 上でどのような

広告活動を行っているか分析するために、アフィリエイト ID を用いてアフィリエイトを特定するアフィリエイト ID 特定分析の提案とアフィリエイト分析支援システムの開発を行った。アフィリエイト ID に着目し、分析することによって複数のブログサイトにまたがるアフィリエイトの分析と掲載商品のジャンル分析が可能となった。今後の課題としては、ブログサイト間のリンク解析、広告掲載位置と本文内容を確認する為に、ブログ記事本文の分析とブログ記事を本文と付随情報部分に分けての分析が必要である。

文 献

- [1] 長谷巧, 山本匠, 原正憲, 山田明, 西垣正勝. アフィリエイトに着目したスパムブログ評価方式に関する検討. 情報処理学会研究報告, 2009-CSEC-44, pp.97-102, 2009.
- [2] Y. M. Wang, M. Ma, Y. Niu, and H. Chen. Spam double-funnel: Connecting web spammers with advertisers. In Proceedings of the 16th WWW Conference, pp.291-300, 2007.
- [3] 齊藤香里, 村井源, 住住彰文. 心の状態と言語の特徴：ブログにおける商品紹介分の分析. 情報知識学会誌, Vol.19, No.2, pp.144-151, 2009.
- [4] 石井聡一, 芳中隆幸, 福原知宏, 増田英孝, 中川裕志. Web 上の広告情報を用いたアフィリエイトスパムの分析. 楽天研究開発シンポジウム 2009, pp.49-52, 2009.
- [5] 伊藤哲哉, 坂巻隆之, 今村仁彦, 鈴木将司. アフィリエイトの達人養成講座. 翔泳社, 2007.