

wwwにおけるクチコミ情報収集の方法に関する考察 ：人の情報探索行動の観点から

メタデータ	言語: Japanese 出版者: 大阪市立大学経営学会 公開日: 2018-06-07 キーワード (Ja): クチコミ, webコミュニティ, ワールド・ワイド・ウェブ キーワード (En): berrypicking 作成者: 竹岡, 志朗, 高木, 修一 メールアドレス: 所属: 桃山学院大学, 富山大学
URL	https://doi.org/10.24544/ocu.20180609-003

Title	www におけるクチコミ情報収集の方法に関する考察：人の情報探索行動の観点から
Author	竹岡, 志朗 / 高木, 修一
Citation	経営研究. 69(1); 91-107
Issue Date	2018-05-31
ISSN	0451-5986
Textversion	Publisher
Publisher	大阪市立大学経営学会
Description	

Osaka City University

www におけるクチコミ情報収集の 方法に関する考察

—— 人の情報探索行動の観点から ——

竹 岡 志 朗・高 木 修 一

目次

- 1 はじめに
- 2 インターネット上の情報源と消費者による情報探索行動
 - 2.1 クチコミの消費者行動に与える影響
 - 2.2 消費者による情報探索
 - 2.3 web コミュニティ
 - 2.4 先行研究の課題
- 3 調査の方法
- 4 調査の結果
 - 4.1 google の検索結果に基づいた調査結果
 - 4.2 web リンクに関する調査結果
- 5 調査の結果の考察
- 6 今後の課題

1 はじめに

インターネットの普及は消費者による購買意思決定過程に大きな変化をもたらし続けており、2017 年には「インスタ映え」という言葉が流行語になるとともに、消費を喚起・促進するキーワードとして頻繁に言及されている。

総務省が行った平成 28 年通信利用動向調査によると「過去 1 年間にインターネットを利用した」割合は 76.4%¹⁾、インターネット利用者の利用頻度は「毎日利用する」が 73.1%である。また古いデータではあるが平成 17 年度の同調査によると、パソコンまたは携帯電話・PHS からのインターネット利用者を対象とした「過去 1 年間にパソコンまたは携帯電話からインターネットを利用した目的・用途」における「商品・サービス・企業・店舗等の情報入手」の割合は 50.5%に上っている²⁾。

インターネットの普及と利用形態の発展と定着、またそれによる社会の変化などを背景に、経営学諸分野においても、消費者や web、それらの関係に関する様々な研究がなされてきた。

キーワード：クチコミ、web コミュニティ、berrypicking、ワールド・ワイド・ウェブ

[受理日 2018 年 2 月 6 日]

これらの研究の多くは、インターネット上のコミュニティ・サイトや掲示板、商品購買レビューに関するものが中心である。これらを研究の素材として利用する利点としては、比較的収集が容易であること、興味関心に沿って集まったユーザーによって書き込まれたテキストであるため分析前の選別が不要であること、フォーマットが統一されているため、分析の前処理が容易であることなどがあげられる。一方でデメリットとしてはこれらのテキストは web 上に多数存在するテキストの一部でしかない、つまりデータに偏りが生じてしまうことがあげられる。

本研究は web 上のテキストデータをより活用するための基礎的研究として、キーワードに基づく境界線の無い web コミュニティの分析可能性について検討する。

2 インターネット上の情報源と消費者による情報探索行動

2.1 クチコミの消費者行動に与える影響

消費者の情報探索に利用されるものとして一对多型の企業による web ページや消費者によって書き込まれ蓄積される web ページ・ブログ、多対多型のコミュニティ・サイトや掲示板がある。この中でも特に後者、つまり消費者によって書き込まれた情報が注目されている。その理由としては、企業による web ページは利用者は多いものの、重視する情報源としてはクチコミ・サイトが上回っている（濱岡, 2006）ためである。

なぜ人はクチコミを発信するのか、この問いに対して斉藤（2015）はクチコミ投稿者がどのような消費者への情報発信を意図しているのかの違いから次のように明らかにしている。当該カテゴリーの商品を未経験の消費者に対して情報を発信する意図を生み出す要因としては、他者に先駆けて使用した経験を印象付ける先進的自己の呈示などがあげられ、他方で経験者への情報発信の意図を生み出す要因としてはその商品への満足度と知覚された新しさがあげられている。

このように発信されるクチコミは消費者の購買意思決定時の重要な情報源となっている。なぜなら、購買前に直接的に品質を確かめることが困難な経験品質³⁾について、他者の経験を通して情報を得ることができるためである。

インターネット上のクチコミの消費者への影響に関する研究は、それまでの face-to-face のクチコミに関する研究を土台として様々な研究が蓄積されてきた。澁谷（2006, 2007）は従来のクチコミとの比較においてインターネット上のクチコミの利点として膨大な情報が検索・収集可能で社会的文脈依存性の限界を克服することができること、欠点としてインターネットの匿名性によって情報の発信者に対する信頼性を構築することができず、クチコミが本来持っていた強みを失ってしまったことを指摘している。この匿名性を補うのがクチコミの送り手と受け手の関連属性における類似性である。澁谷（2007）は消費者が経験品質の評価に関して探索する際には、自身と類似する属性を持つ他者の発信したクチコミ情報により強い影響を受けることを明らかにしている⁴⁾。

Vázquez-Casielles et al. (2013) は消費者の製品（携帯電話とノート PC）とサービス（携帯電話のキャリアと旅行代理店）についてのクチコミと消費者の購買行動の関係を分析している。Vázquez-Casielles et al. は肯定的なクチコミも否定的なクチコミも消費者の購買可能性に影響を与えるが肯定的なクチコミの方がより強いこと、クチコミの送り手の表現の強さが肯定的なクチコミであれ、否定的なクチコミであれ、影響力が強いこと、事前のブランド・ロイヤリティがクチコミの購買への影響を減少させることなどを明らかにしている。

また、Berger & Milkman (2012) は SNS の登場によって関心の共有やバイラル・マーケティングが促進されているという背景の中で、クチコミに関する研究がその効果にばかり注目し、何がクチコミを共有することを促進するのかに関しては十分に明らかになっていないことを指摘した。そのうえで、オンライン上のコンテンツは、それが肯定的なものであれ否定的なものであれ感情を強く刺激するものほど共有され広がっていくことを明らかにした。

このようにインターネット上におけるクチコミは消費者の意思決定に大きな影響を与えていることが明らかとなっている。では、消費者はどのようにしてクチコミなどの情報をインターネットから取得しているのだろうか。

2.2 消費者による情報探索

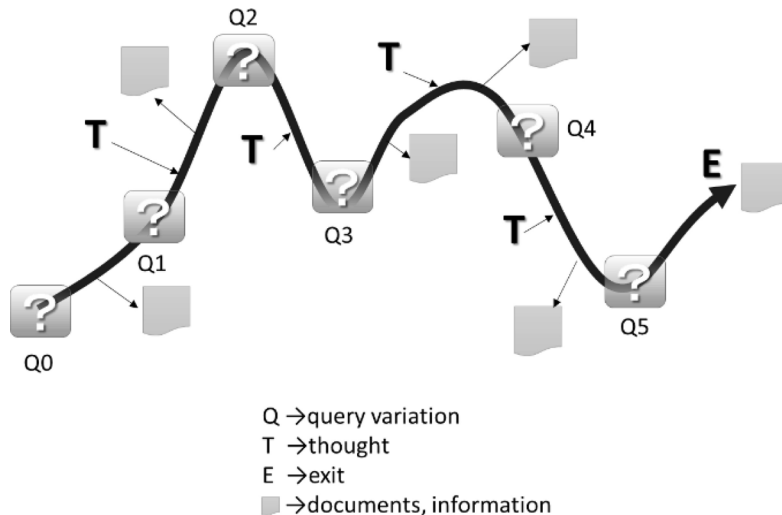
インターネットは TV 広告などとは異なり、能動的に情報を探索しなければならないメディアである。そのため消費者はクチコミも含めた様々な情報を能動的・積極的に探索し、学習する必要がある⁵⁾。

人の情報探索行動は新しい情報を連続的に発見する過程であると同時にその発見した情報を学習する過程でもある。情報を収集する過程に関するモデルとしては Bates (1989) の berrypicking モデル（図 1）がある。

berrypicking モデルの特徴は、それまでの情報探索が情報へのニーズから生み出されるクエリーとその結果得られるドキュメントの一致だけを論じていたのに対して、人の情報探索過程における探索した情報からのフィードバックとそれによる思考と学習、探索過程の中での柔軟なクエリーの再設定なども取り込んだ点にある。

この berrypicking モデルは検索エンジンを用いた web における情報探索行動についてのものではない。しかし、江草ら (2009) の研究で web における情報探索行動においても berrypickikng モデルと同様の情報探索行動が行われていることが示されている。江草ら (2009) は学部生と大学院生を対象に情報探索行動に関する調査を行った。この中で被験者として登場する web ブラウジングに慣れた大学院生の情報探索行動は、最初の検索結果から web ページの閲覧、リンク先 web ページの閲覧、検索結果に戻り他の web ページの閲覧、検索ワードの変更などの過程で進められており、この研究の中では明示されていないが berrypicking モデルと同様の情報探索行動が行われているといえる。また、インターネットを用いた情報探

図1 berrypicking モデル



Bates (1989) より引用

索に慣れている者ほど link depth^{2 6)}以上の web ページを link depth1 の web ページよりも閲覧していることも示されている。

また、寺井（2011）の大学生を対象にした情報探索行動に関する研究の中では、探索対象の明確さが web 探索における新規キーワードの生成に正の影響を及ぼしていることが示されている。同様に探索内容の明確さが探索行動に与える影響について研究した齊藤・橋本（2006）では、探索内容が明確な課題を与えられた被験者は、探索内容が不明確な課題を与えられた被験者と比べてより検索結果から辿るリンクが深い、つまり web ページ上のリンクをより活用する傾向があることが示されている⁷⁾。

このような情報探索における認知的情報処理、つまりキーワードの設定や発見された情報の学習、さらなる検索への反映、つまり berrypicking の認知的過程について論じているのが齋藤・三輪（2003）と齋藤（2006）である。これら研究では情報探索に Simon & Lea（1974）の two space モデルを援用することで情報探索における認知的情報処理の過程を明らかにしようとしている⁸⁾。これらの研究は web 上の情報探索行動にキーワードや検索式を考える段階と検索結果として示されたページのリストや個々のページを探索する段階があり、これらが循環するサイクルとして情報探索が行われていると考えている。前者の段階の探索はキーワード空間と呼ばれる人の内的空間で、後者の段階の探索は Web 空間と呼ばれる外的空間で行われる。これらの研究では www（world wide web）に関する知識や経験から被験者を Expert と Novice に分け、それぞれの検索行動の違いを two space モデルから説明している。この中では Expert と Novice の間にキーワード空間の広がりには差はない、つまりキーワードとして用

いられる語には差がないこと、しかしキーワードの有効性やその検索結果に関する予測については Expert がより行っていたことが示されている。また Expert は検索結果の上位を集中的に探索していること、探索結果から得られた情報をキーワード空間によりフィードバックしていることが明らかにされている。

2.3 web コミュニティ

インターネット上には様々なクチコミが存在し、我々の情報探索行動はそれらを berrypicking することで進められる。このような消費者が探索するクチコミが交換・蓄積されているのが @コスメのようなコミュニティ・サイトや価格コムにある商品ごとの掲示板、SNS 上のコミュニティである。

齋木（2002）はコミュニティ・サイトについて、その強みとしてコミュニケーションが活発に行われることで、そのコミュニケーションに人が集まり、また新たなコミュニケーションが生まれること、そのコミュニケーションプロセスから参加者の問題解決となる情報が生まれ、それによって参加者のコミュニティへの信頼感が高まり、そのコミュニティ内でのコミュニケーションを目的に、さらなる参加者が集まることを指摘している。また、石井（2002）は @コスメのようなコミュニティ・サイトが新たな利益の源泉やビジネスチャンスになりうるとし、成功するコミュニティ・サイトについて 7 つの観点から論じている。この中では、通常コミュニティを考える場合、参加者の継続的な参加が重視されるが、コミュニティ・サイトの場合には継続するコミュニケーションこそがコミュニティ・サイトの要件だと指摘している。

また、吉見（2013）はクチコミが投稿されるプラットフォームの違い、具体的には Amazon と楽天の間で同一の商品に関するクチコミや評価であっても、そこで使用される単語に明確な違いがあることを明らかにしている。

以上のようなインターネット上のクチコミに関する研究の多くはコミュニティ・サイトや掲示板、通販サイトに投稿されたレビューを分析の対象として蓄積されてきた。しかし、web 上にある情報はこのようなまとまったものだけではない。検索エンジンにキーワードを入力し berrypicking を行えば数多くの web ページやブログ記事を発見し、閲覧、知識を蓄積することが可能である。これら検索エンジンによって集められた web ページやブログ記事は境界線の無い、しかし同一の興味・関心を持った web ページ群といえる。

web 構造分析に関する諸研究では、これら web ページ群を web コミュニティ⁹⁾として析出する方法が提案されている。Kumar et al. (1999) は web ページ間のリンクを分析することで発見されるという完全 2 部グラフを web コミュニティのコアとし、このコアを中心に web コミュニティを発見することができるとしている。このようなコミュニティの発見手法は、無数に存在する web ページ作成者の興味・関心に沿って作成された web リンクの共引用関係からコミュニティを析出することから、その発見されたコミュニティ内に属する web ページの作成

者が想定しなかったような形の web コミュニティが発見される可能性もある。

村田 (2001) も Kumar et al. (1999) と同様にハイパーリンクにおいて析出される完全 2 部グラフを web コミュニティとしている。Kumar et al. (1999) は web クローラーを用いて web リンクを抽出し、その結果に基づいてコミュニティを析出していたが、村田 (2001) はより簡便な方法として検索エンジンを用いた完全 2 部グラフに基づく web コミュニティの発見手法について述べている。この方法では、まず URL をあらかじめいくつか決めておき、その URL を検索エンジンを用いて検索、その URL にリンクを張っている web ページを発見し、それを fans とする。次に fans の URL にアクセスし、それぞれの web からハイパーリンクを抽出、その中で最も参照数の多い URL は fans が多く参照している URL であり、web コミュニティの centers ということになる。このような処理を繰り返すことで村田は完全 2 部グラフに基づく web コミュニティを発見できるとしている。

また、久我と中所 (2001) は web サイトのランキングを作成するという目的のもと、リンク集に基づく web コミュニティの発見手法について述べている。この中ではリンク集はあるテーマに精通した人が作成しているということ、リンク集はひとつでは十分な意味を持たないので複数のリンク集を集めること、複数のリンク集からリンクされている web ページはそのテーマに精通している人がリンクしていることから、web ページのランクが上位であると推測できることを指摘している。そして、あるテーマに基づくリンク集を持つサイトの集合を web コミュニティと呼んでいる。

以上のように、コミュニティ・サイトだけではなく、web リンクに基づいて境界線の無い web コミュニティを探索する手法が提案されている。次節以降ではこのような web のリンクに着目して、境界線の無い web コミュニティにおけるクチコミ研究の分析可能性について調査を行っていく。

2.4 先行研究の課題

web コミュニティに関する研究は web ページの内容には触れず、web ページ間の関係が作り出すグラフ構造やリンク関係に注目し、web コミュニティを析出しようとしていた。しかし、情報探索はキーワードに基づいて行われる探索行動であることから、web ページの中身、特にキーワードに注目しないわけにはいかない。以下ではキーワードを中心にリンク関係に基づく web コミュニティを取り出すことが可能なのかを検討する

3 調査の方法

これまでの web コミュニティ研究の課題を克服するために、本稿では web コミュニティを「共通の話題について記載されたハイパーリンクによってつながり、形成される web ページ群」とする。web 構造に基づく web コミュニティの析出の方法では、そのコミュニティが実際に

どのような話題についてのものかは十分に検討できていなかった。本稿ではキーワードを中心に web コミュニティを析出することで、web 構造に基づく研究の課題を克服するとともに、web クローラーによる探索によって人の情報探索行動と同型の探索行動を実現する。キーワードが含まれている web ページをノード、web リンクをリンクとしたネットワークを析出することができれば、それが web コミュニティといえる。

www と web リンクに関する理想の状態としては、キーワードを有する web ページを初期値としてひとつ決定することができれば、その web ページからはじめて web リンクをたどることで、すべての web ページに到達することができる。しかし、今回は、人の情報探索行動の模倣によるコミュニティ析出の可能性の検討ではなく、分析対象としての web コミュニティの析出の可能性の検討が目的であるため、情報探索の初期値として google の検索結果を用い、クローリングのスピードを上げることとする。

今回はキーワードとして Mojang 社が開発したサンドボックスゲーム「Minecraft」を選択した。世界的に人気があり、日本でも小学生から大人まで多様で多数のプレイヤーが存在し、また Education Edition といった形で学校における授業の中で使われることが、その理由である。また、ナンバリングタイトルが存在しないため、モノと単語が一意に決まるという分析上の利便性も考慮している。

データの収集には web クローラーを用いる

1. google 検索でクローラーが最初に探索する URL のリストを取得
2. クローラーを用いてリストにある URL を探索し、web ページにキーワードが含まれているのかを確認
3. 取得した URL をデータベースに新たに登録するとともに、その web ページの情報もデータベースに登録（キーワードが無くても web ページの情報と URL は取得し登録する）

google を用いて「Minecraft」、「マイクラフト」、「マイクラ」を検索した結果 316 件の URL（重複削除済み）を取得することができた。さらにクローリングの速度を上げるために google サジェストの機能も活用し、1952 のキーワードを新たに追加で検索（重複削除済み）、240058 件の新たな URL を取得（重複削除済み）することができた。

表 1 google での検索の結果¹⁰⁾

キーワード	ヒット件数	実際の件数	サジェスト
Minecraft	55件	66件	1053種類
マイクラフト	9490000件	162件	1064種類
マイクラ	12600000件	153件	1056種類

出所：筆者作成

google による初期 URL を取得する期間を含めたデータの収集期間は 2017 年 9 月 3 日から 11 月 3 日である。

クローラーによる web ページの巡回では、キーワードの有無と web リンクの確認を行っているが、日本語の web ページからリンクされるものだけを分析の対象に（言語設定がされていない web ページは含む）、javascript によるリンクはたどらない、動的 URL は ? や + などの記号以下を削除し登録するといった処理を行っている。

クローリングは google の検索結果として取得できた web ページを第 1 層とすると、第 2 層までを対象とし、第 3 層は 2 % の web ページをサンプリング調査した。

4 調査の結果

4.1 google の検索結果に基づいた調査結果

村田（2001）では、検索エンジンの結果を用いて完全 2 部グラフを作成し web コミュニティを析出していた。今回の調査では google の検索結果をクローリングの開始 URL として用いているが、google を活用するだけで web ページを取得することができればより時間の短縮が可能である。

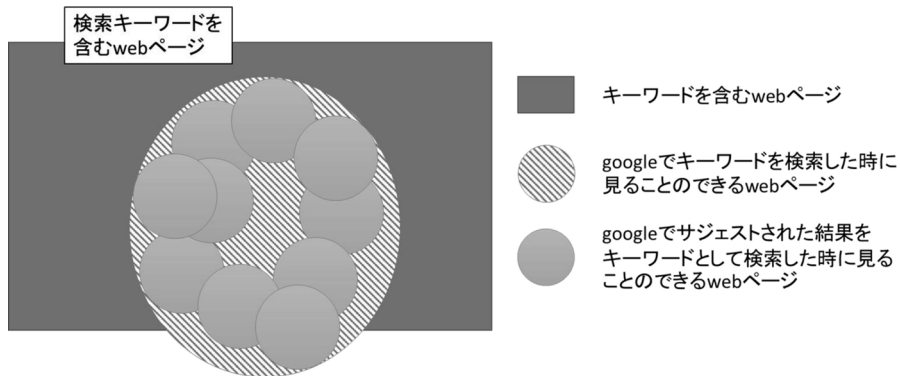
表 2 取得した web ページ

	Webページ数	Webページ数(キーワード有無別)	検索適合率
第1層	160444(日本語) (全240374)	あり	124803 77.8%
		なし	35641 googleの検索結果
第2層	3211253(日本語)	あり	1115443 53.2%
		なし	2095810
第3層	56090637中 1091695を確認	あり	182758 16.7%
		なし	908937 (2%のサンプリング調査)

出所：筆者作成

表 2 は各層で取得することのできた web ページの数と、それら web ページのキーワードの有無を集計したものである。google の検索結果からアクセスした第 1 層には日本語の web ページが 160444 ページ、うちキーワードの存在したページが 124803 ページで、キーワードの存在した割合は 77.8% である。この結果は google で検索を行い URL を取得した日と実際にクローラーによって確認した日の間にずれがあったことが原因と考えられる。また第 2 層、つまり google の検索結果から得られた web ページに記載のあった web リンクをたどって取得した web ページにもキーワードの存在する web ページが 1115443 ページあった。このことから図 2 のように google を用いて検索しただけではたどり着くことの困難な web ページが相当数存在するといえる。

図2 www と検索結果



出所：筆者作成

4.2 web リンクに関する調査結果

今回の調査では初期データとして google の検索結果を用いたが、google の検索結果を用いることができなければ web リンクだけを用いて web コミュニティを析出する必要がある。すべての web ページが web リンクのものであり web リンクの先になりうるため、すべての web ページの参照・被参照関係を明らかにする必要がある。

表3は今回取得することのできた 3371697 ページの中に含まれている第1層から第2層、第1層から第1層、第2層から第1層、第2層から第2層への web リンクを集計したもので、どのような web ページからどのような web ページにリンクしているのかをまとめたものである。

表3 キーワード・ドメイン別リンク数

リンク元キーあり	リンク元キー無し	リンク数	ドメイン	リンク数	割合
あり	あり	338046448	同一ドメイン	335070950	58.6%
			ドメイン外	2975498	0.5%
なし	あり	4457822	同一ドメイン	3786393	0.7%
			ドメイン外	671429	0.1%
あり	なし	33053938	同一ドメイン	19664012	3.4%
			ドメイン外	13389926	2.3%
なし	なし	195841126	同一ドメイン	175887876	30.8%
			ドメイン外	19953250	3.5%
計		571399334 (取得：585967002)	同一ドメイン計	534409231	ドメイン内率：93.5%
			ドメイン外計	36990103	ドメイン外率：6.5%

出所：筆者作成

取得できた web リンク 585967002 件、うちキーワードの有無を確認できた web リンクは 571399334 件、1web ページあたりの平均リンク数は 201 であった。この結果から、下記のこ

とが言える。

1. 同一ドメインへのリンクが全体の 93.5%、ドメイン外の web ページへのリンクは 6.5% ということから、現在の web サイトが、かなりの程度、閉鎖的に運営されていることが推測される
2. キーワードを持つドメイン外の web ページへのリンクは 0.6% で、web クローラーはこの 0.6% しかないリンクをたどることで他のドメインの web ページを探索の対象にすることができる（人の情報探索行動も同様のことが言える）
3. キーワードを含む web ページからキーワードを含む web ページへのリンクは全体の 59.2%（四捨五入あり）
4. キーワードを含まない web ページからキーワードを含む web ページへのリンクも存在する、（全体の 0.8%（四捨五入あり））

また、キーワードの存在するドメイン外へのリンクの詳細を見ると、リンク数 3646927 件の中で、被リンク数 100 位までの web ページで全体の被リンクの 32.9%（被リンク数：1201463 件）となっており、外部ドメインへの web リンクも偏ったものとなっていることがわかる。（表 4）

表 4 リンク先上位 10 件の URL

URL	被リンク数	内容
http://3d.nicovideo.jp/	89924	3d画像共有ポータル
http://com.nicovideo.jp/	86878	動画サイトポータル
http://ch.nicovideo.jp/	79814	動画サイトポータル
http://ch.nicovideo.jp/portal/blomaga	77706	ブログポータル
https://www.youtube.com/watch	43124	動画サイトポータル
http://b.hatena.ne.jp/entry/	40019	ブログポータル
https://chiebukuro.yahoo.co.jp/dir/dir_list.php	29646	質問投稿サイトポータル
https://chiebukuro.yahoo.co.jp/ranking/ranking.php	28201	質問投稿サイトランキング
http://www.bookdepository.com/	21870	通販ポータル
https://chiebukuro.yahoo.co.jp/dir/list/d2078297513/solved	21769	質問投稿サイト検索結果

出所：筆者作成

最後に、表 5 にあるように、ドメイン別の被リンク数を集計すると、今回取得できた全 151805 ドメイン中 9140 ドメインが外部ドメインからの被 web リンク数が 10 件以下であり、被リンク数が 1 件のドメインも 2865（1.9%）あることが分かった。これらのドメインは人の情報探索行動においても、クローラーを用いた探索行動においても、キーワードに基づいて探索した場合、1 ルートしかアクセスすることの可能なルートが存在しないことを意味している。

表5 被リンク数別トメイン数

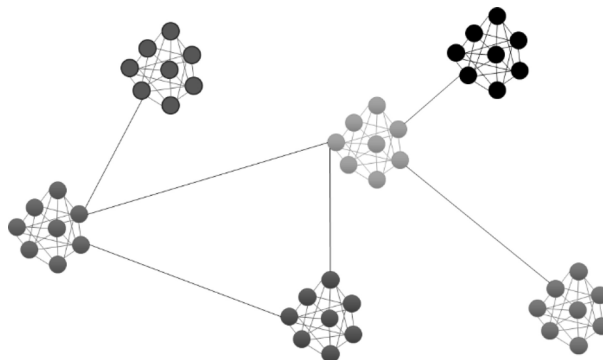
被リンク回数	ドメイン数
1	2865
2	1600
3	1200
4	876
5	675
6	564
7	404
8	393
9	306
10	257

出所：筆者作成

以上から、現在の www の構造について下記（図3はイメージ図）のことが言える。

- 密な web サイト内リンク
 - 同一ドメインへのリンクが全体の 93.5%
- 疎な web サイト間リンク
 - ドメイン外の web ページへのリンクは 6.5%
 - ドメイン外の web ページへのリンクの中でキーワードありページへのリンクは全体の 0.6%
 - 上記リンクの中で情報の広がりの無い（ポータルやランキングの結果、検索結果）リンクは少なくとも 25.1%
 - また 1.9%のドメインが1ルートしかたどり着くルートの無いドメイン

図3 www イメージ図



出所：筆者作成

5 調査の結果の考察

本研究では「キーワードを中心にリンク関係に基づく web コミュニティを析出することは可能か?」という問いに関して調査を行い、下記の結果を得ることができた。

① キーワードを含む web ページを検索エンジンだけを頼りに見つけることは困難

google で得ることのできた web ページは 160444 ページ (キーワードありは 124803 ページ)、リンクをたどって見つけることのできたキーワードの存在する web ページは 1115443 ページ。今回確認することのできた web ページ (1275887 ページ) の中で、google を用いることでたどりつけた web ページは 12.6% であり、google サジェストだけではなく、すべての可能性のある語を組み合わせで検索しても困難だと考えられる。

② www 全体をクローリングできればキーワードを含む web ページを取り出し、それらをコミュニティとして定義することもできるが、困難

今回の調査では 120 万ページ/日のペースでクローリングを行ったが、www 全体で 1 兆ページ以上 (2008 年 google 調べ) あるとされていることから全 web ページをクローリングするには少なくとも 2283 年かかる。また、仮に第 3 層までに限定しても 59462334 ページあり 50 日以上必要となる。今回は第 2 層までに限定し探索を行ったが、探索 web ページは第 3 層までの web ページ数の 6% でしかないにもかかわらず、データベースのテーブル構成を蓄積用に最適化したものから分析用に最適化する処理に 4 日必要だった。以上から、企業のマーケティング部門や経営学の研究者には困難と考えられる。

③ キーワードの存在する web ページからたどることのできる web ページだけを取得するように設定することでの web コミュニティの析出は困難

キーワードを含まない web ページからキーワードを含む web ページへのリンクが 4457822 件存在した。

④ 外部ドメインへの web リンクが少数しか存在せず、そのためクローリングによる web コミュニティの析出は困難

早い段階で関係するドメインを数多く見つけることができれば、クローラーの台数を増やすことができる (同一ドメインへの連続アクセスを回避できる) ためデータ収集のスピードを上げることができるが、キーワードを持つドメイン外へのリンク数が全リンク中 0.6% ではスピードを上げることは困難である。また、1 ルートしかたどり着くルートのない可能性のあるドメインが今回の調査では 2865 (取得できたドメインの 1.9%) あり、web ク

ローリングで全体をくまなく網羅することは困難と考えられる。

以上から、本研究ではキーワードを中心にリンク関係に基づく境界線の無い web コミュニティを取り出すことは困難であり、クチコミ分析を行う場合には、何らかのコミュニティ・サイトにまとめて存在するデータを用いる方法が現実的という結論を得ることができた。

また、今回の調査では初期 URL に google の検索結果を用いて 240374 件準備した（3 日強）ため高速にデータを収集することができたが、この方法では多数の機種をまたいだクチコミの収集（過去にデジタルカメラ 160 機種（竹岡他, 2014）や携帯電話 159 機種（竹岡他, 2015）を分析する事例として用いた）は大変困難であることが分かった。

6 今後の課題

今回キーワードとして選択したのは「Minecraft」であるが、近年のゲームに関する情報共有は動画を通じて行われていることが多い。しかし、本研究での分析手法はテキストデータを収集することが前提となったものである。今後の課題としては、動画を用いたクチコミや情報共有についても分析する方法についても検討する必要がある。また、動画のコメント欄を分析するのであれば、現在の方法でも可能だが、アクセス先が限定されるため、これについても十分な検討が必要である。

謝辞

本研究は科学研究費補助金（日本学術振興会）若手研究（B）・課題番号 17K13787 平成 29～30 年度（研究代表者：竹岡志朗「イノベーションの普及過程で選考される意味属性のテキストマイニングによる可視化」）の助成のもとに行っているイノベーションの普及研究の一環としてなされたものです。

注

- 1) 「いいえ」と回答した割合は 15.1%、無回答は 8.4%である。
- 2) この質問項目は平成 18 年度以降あがっておらず、最新の平成 28 年度調査の中で類似する項目に、SNS の利用目的として「従来からの知人とのコミュニケーションのため」の 82.8%に続いて「知りたいことについて情報を探すため」の 43.7%がある。
- 3) Zeithaml (1981) は様々な商品やサービスの品質を評価の容易なものから困難なものという基準で並べ、評価の容易なものから探索品質、経験品質、信頼品質と 3 分類している。最も評価の容易な探索品質とは消費者が購入前から評価可能なもので服や宝石などがこれに当たり、経験品質とは購入後に使用の中で判断できるものでレストランにおける食事やヘアカットがこれに当たる。信頼品質は消費したのちにもその評価が不明瞭なこともあるもので医療や法律に関するサービスがこれに当たる。
- 4) 澁谷 (2006, 2007) では経験品質ではなく経験属性という言葉を使用している。澁谷 (2006, 2007) が Zeithaml (1981) の experience qualities を引用する中で経験品質ではなく経験属性として用いてい

る点については確認したうえで、本稿でも Zeithaml (1981) を引用していることから、澁谷 (2006、2007) の当該部分については経験品質と言い換えている。

- 5) どのような消費者がよりインターネットを活用し、情報探索を行っているのかという研究があるが、これらの研究結果は異なっている。たとえば、商品に関してあらかじめ持っている情報量の差がインターネットにおける情報探索行動に差をもたらすという仮説に関する研究では大田 (2014) ではより製品について知っている人ほどクチコミ情報を参照しているとしているが、中川 (2013) では製品についてあまり知らない消費者ほど、人的情報源以上にインターネット上の情報源を積極的に活用しているとしている。前者が紙おむつに関する情報探索について、後者が旅行に関する情報探索について調査した結果から導出された結論であることから、探索対象によって異なっている可能性がある。また、商品の属性によって重視される情報源という観点からの研究として清水 (2004) がある。この中では食料品から日用品、サービスまで様々なクチコミの対象を 16 分類し、これらを利用する上での情報源として何が重要視されているのかについて分析を行っている。たとえばアルコール飲料の場合には「店頭で見る」や「テレビコマーシャル」、「チラシ広告」などが上位の情報源としてあげられるが、自動車の場合には「店頭で見る」が圧倒的に多く、比較検討の段階から最終決定の段階に進むにつれて「店頭で聞く」という情報源の重要性が高まることが明らかにされている。また、積極的に活用される情報源とともに、どのようにその情報源が活用されるかについても明らかにされている。たとえば、家電や自動車の場合には情報処理型（消費者が自らの目標を定め、積極的に情報を収集する）の外部情報の探索が、ソフトドリンクやアルコール飲料の場合には刺激－反応型の外部情報利用がなされていることが示されている。
- 6) 江草らは検索エンジンを用いた検索結果から辿って閲覧した web ページを link depth1、link depth1 の中にあるリンクを辿って閲覧した web ページを link depth2 と呼んでいる。
- 7) 検索結果の一覧から閲覧される web ページの数には有意な差はない。
- 8) このようなサイクルを学習としてとらえた研究として吉岡 (2007) がある。吉岡 (2007) は、インターネットによる検索が、検索者があらかじめ持っている知識と、検索した結果から得られる知識を関連付けて、目的に向けて新たな知識を構築していく過程であり、認知心理学における知識構築過程と同様のものとし、このような学習過程を支援する方法について論じている。この中ではインターネット上での情報検索と知識構築を有効なものとしているのが探索者のメタ認知、つまり検索テーマについての既有知識や検索経験を検索のリソースにするメタ認知と web 上の文章から有効な情報を得て検索のリソースにするメタ認知である。これらのメタ認知を活性させることで学習のサイクルが回ることから、メタ認知を促す支援を行うことで、探索者の知識構築を支援することができるとしている。
- 9) web コミュニティという表現は web 構造に基づいてコミュニティを析出しようとする研究群の用法を採用している。例えば村田 (2002) は「同一のトピックに関する Web ページ集合」という意味で用いており、本稿でもこの意味で、つまり「Minecraft」というトピックに関する web ページ群という意味で用いている。このようなコミュニティという概念の使用は誤解を招くことになりかねない。例えば、石井 (2002) ではコミュニティの定義について「共有された価値」がカギとなるとしている。しかし、インターネット上の同一のトピックに関する web ページ群に共有された価値は存在するだろうか。共通の話題についてハイパーリンクによってつながった群という意味では何らかの価値の共有は存在しているだろうが、従来のコミュニティ概念とは異なるものといえる。このような差異は、コミュニティの主体に人を置くのか、それともモノを置くのかにある。通常のコミュニティ概念の中心は人であり、人が価値を共有し人がコミュニケーションなどでつながって形成されたものがコミュニティとされているが、web コミュニティ概念は web ページに記載されたものに共通点があり、web ページがハイパーリ

ンクによってつながって形成されたものがコミュニティとされているのである。本稿では、web 構造研究を土台として研究を進めるためコミュニティ概念について web 構造研究の定義を採用しているが、その意味するところは「共通の話題について記載されたハイパーリングによってつながり、形成される web ページ群」である。

- 10) google で検索を行うと検索結果の一覧ページに何件の web ページが見つかったのかを提示している。しかし、実際にその検索結果からたどり着くことの可能な web ページは少ない。たとえば、「マイクラフト」を google で検索すると 9490000 件の web ページが見つかるが、実際に web ページとして検索結果からたどり着くことのできる web ページは 162 件であった。また、「Minecraft」のヒット件数が他と比べて少ないが、これは筆者が確認の過程で 1 度すべてのリンク先を検索結果から確認した後に、再度ヒット件数を確認するために検索を行ったため、このような少ない件数になった。

参考文献

- 総務省 (2006) 『平成 17 年通信利用動向調査』。
- 総務省 (2017) 『平成 28 年通信利用動向調査』。
- Bates, M. J. (1989). "The design of browsing and berrypicking techniques for the online search interface," *Online review*, vol. 13 (5), pp. 407-42, (<https://pages.gseis.ucla.edu/faculty/bates/berrypicking.html>).
- Berger, J., & Milkman, K. L. (2012). "What makes online content viral," *Journal of marketing Research*, vol. 49 (2), 192-205.
- Kumar, R., Raghavan, P., Rajagopalan, S., & Tomkins, A. (1999) "Trawling the Web for emerging cyber-communities," *Computer networks*, vol. 31(11), pp. 1481-1493.
- Simon, H. A., & Lea, G. (1974). "Problem solving and rule induction: A unified view," in Gregg L., W., (eds) *Knowledge and cognition*, pp. 105-127, Lawrence Erlbaum.
- Vázquez-Casielles, Rodolfo, Leticia Suárez-Álvarez, and Ana-Belén del Río-Lanza. (2013) "The word of mouth dynamic: How positive (and negative) WOM drives purchase probability," *Journal of Advertising Research*, vol 53.1 pp. 43-60.
- Zeithaml, V. A. (1981). "How Consumer Evaluation Processes Differ for Products and Services," (https://www.researchgate.net/publication/279912843_How_Consumer_Evaluation_Processes_Differ_for_Products_and_Services, 2017 年 12 月 07 日確認)。
- 石井淳蔵 (2002) 「コミュニティとコミュニティ・サイトの理論的基礎」石井淳蔵・厚美尚武編『ネット・コミュニティのデザイン インターネット社会のマーケティング』有斐閣。
- 江草由佳, 高久雅生, 齋藤ひとみ, 寺井 仁, 三輪眞木子, 神門典子 (2009) 「Link Depth: Web 情報探索行動の閲覧パターンの分析」『研究報告情報学基礎 (FI)』no. 20, pp. 1-7.
- 大田謙一郎 (2014) 「消費者知識とメディア選択行動」『長崎県立大学経済学部論集』vol. 48(3), pp. 81-96.
- 久我昌崇, 中所武司 (2001) 「Web コミュニティの知識に基づく情報検索手法の評価」『情報処理学会研究報告データベースシステム (DBS)』pp. 37-44.
- 齋木 (2002) 「アクティブなコミュニティ・サイト 妊娠・出産・子育てを応援『ぶれままクラブ』」石井淳蔵・厚美尚武編『ネット・コミュニティのデザイン インターネット社会のマーケティング』有斐閣。
- 斉藤嘉一 (2015) 『ネットワークと消費者行動』千倉書房。

- 齋藤ひとみ (2006) 「Web 情報探索における認知プロセスの理解とその応用」『情報知識学会誌』 vol. 16(4), pp. 51-62.
- 齋藤ひとみ, 橋本恵美子 (2006) 「Web 情報探索におけるブラウジングプロセスの分析: 探索内容の明確さの違いによる検討」『日本教育工学会論文誌 30 (Suppl.)』, pp. 57-60.
- 齋藤ひとみ, 三輪和久 (2003) 「問題解決活動としての WWW 情報探索: 科学的発見の枠組みに基づく検討」『認知科学』 vol. 10(2), pp. 258-275.
- 澁谷 覚 (2006) 「インターネット上の情報探索: 消費者によって発信された体験・評価情報の探索プロセス」『消費者行動研究』 vol. 13(1), pp. 1-29.
- 澁谷 覚 (2007) 「ネット上の消費者情報探索とネット・クチコミのマーケティング利用」『AD STUDIES』 vol. 20, pp. 11-15.
- 清水 聡 (2004) 「消費者の情報処理プロセスと外部情報の研究」『三田商学研究』 vol. 47(3), pp. 113-127.
- 竹岡志朗, 高木修一, 井上祐輔 (2014) 「テキストマイニングを用いたイノベーションの普及分析」『日本情報経営学会誌』 Vol. 35-1, pp. 72-86.
- 竹岡志朗, 高柳直弥, 井上祐輔, 高木修一 (2015) 「イノベーションは消費者にどのように認知されているのかクチコミ情報から見るイノベーションの非連続性と連続性」『NEXTCOM』 Vol. 21, pp. 54-63.
- 寺井 仁 (2011) 「ハイブリッドな情報環境における情報探索行動に関する実証的研究」『日本図書館情報学会誌』 vol. 57(2), pp. 43-62.
- 中川正悦郎 (2013) 「情報源としてのインターネットと消費者情報源選好」『マーケティングジャーナル』 vol. 33(2), pp. 80-93.
- 濱岡 豊 (2006) 「消費者間の相互作用 クチコミを中心に」 田中 洋・清水 聡編『消費者・コミュニケーション戦略 現代のマーケティング戦略④』 pp. 57-93.
- 村田剛史 (2001) 「参照の共起性に基づく Web コミュニティの発見」『人工知能学会論文誌』 vol. 16(3), pp. 316-323.
- 村田剛史 (2002) 「ハイパーリンクのグラフ構造に基づく Web コミュニティの洗練」『人工知能学会論文誌』 vol. 17(3), pp. 322-329.
- 吉岡敦子 (2007) 「インターネット情報検索における知識構築を促進させるメタ認知の検討」『日本教育工学会論文誌』 vol. 32(2), pp. 115-123.
- 吉見憲二 (2013) 「プラットフォームの違いがクチコミに与える影響に関する研究」日本情報経営学会誌, vol. 33(3), pp. 109-120.

A study of methods for collecting word of mouth information on world wide web

Shiro Takeoka and Shuichi Takagi

Summary

An increasing amount of research is being conducted using online word-of-mouth sources. Most of this research uses data from web communities, message boards, and other sources with a defined scope, thereby analyzing only a portion of web-based, word-of-mouth data. Inherent in this is the problem of data bias. This paper considers information search activities, specifically keyword searches, and investigates the potential of keyword-based, web community analyses. Analyzing 3,371,697 Japanese-language web pages and 585,967,002 links within these pages that are related to a specific product, the analysis yields four different points to illustrate the difficulty in segmenting web communities based on keywords.