

Web マイニングツールを用いた ジェンダー関連 Web コミュニティの通時的分析

Diachronic Analysis of Gender-related Web Communities using a Web Mining Tool

小山 直子[▼] 増永 良文[▲]

Naoko OYAMA Yoshifumi MASUNAGA

現代は、さまざまな組織体や個人が Web サイトを立ち上げて情報発信している時代である。ジェンダーコミュニティも例外ではなく、男女共同参画社会の実現に向けて、国内外に多数のジェンダー関連 Web サイトを立ち上げている。Web ページは HTML で記述され、したがって Web はページ間に非線形に入り組んだ複雑なリンク構造を有しているが、これをうまく解析すると Web コミュニティを発見することができるのではないかという期待から、これまでに多くの Web リンク解析アルゴリズムが提案されてきた。本研究で使用している Companion^{*} もそのひとつである。しかしながら、我々は、新たにリンク解析アルゴリズムの開発に挑戦するのではなく、有望と考えられる既存の Web マイニングツールを用いて、実際に Web アーカイブスを徹底的にマイニングして、どのような知見が得られるのかという研究に取り組んだ。これは、データベースにおけるコンテンツ解析の重要性を強調するもので、これまでややもすればアルゴリズムやシステム開発に重点がおかれてきたデータベース研究に新たな可能性を提示する狙いがある。もちろん、副効果として、とすれば机上の空論に陥りがちなツールの有用性に関して、実のある議論を展開することが可能となったとも考えられる。もちろん、マイニングの対象となった応用ドメインでは、ツールを適用したことで初めて明らかとなる重要な現象の発見も期待できる。「ジェンダー」は、社会的・文化的所産であり時々刻々変化し続け、同時にパラダイム変換を起こす影響力をもつ概念である。どのような現象と結びつき、自らが変化をおこし、他者に対していかなる変化を誘引しむるかという発展性を持つ。通時的変化を観察する目的をもった Web マイニングの応用ドメインとして、内容的にも規模的にも有効であった。

At present a variety of organizations and personals tend to disseminate information via Web pages. Gender community is not an exception so that we can find a lot of gender-related Web sites towards the realization of the gender-equal society. Since Web pages are described by HTML, a complex non-linear link structure is induced on the Web. Many researches have been conducted to develop efficient Web link analysis algorithms so that meaningful Web communities can be discovered.

[▼] 正会員 お茶の水女子大学COE客員研究員
oyama@cc.ocha.ac.jp

[▲] 正会員 お茶の水女子大学理学部情報科学科
masunaga@is.ocha.ac.jp

Companion is such a promising tool that we used in this research. Instead of developing such a tool newly, we changed our viewpoint to mine the Web archives be through in order to find new knowledge in a specific application domain. Of course, an examination of the mining capability of a given tool can be performed deeply which usually tends to fall into an impracticable theory. "Gender" was really a one of the most suitable Web mining application domains for such purpose because it changes its figure diachronically with the times which due to its intrinsic social and cultural characteristics.

1. はじめに

我が国は男女共同参画社会の実現に向けて大きく動き出しているが、それに呼応して社会のジェンダー意識が急速に高まっている。また1960年代に始まったジェンダー研究も時代の流れに呼応して、その姿をめまぐるしく変えながら発展している[1]。我々は、このような現象がサイバースペース(=インターネット)におけるジェンダー関連Webサイトがおりなすコミュニティを分析することにより、的確に捉えることができるのではないかと考えた。そこで、本研究では、東京大学生産技術研究所喜連川研究室で収集された過去5年分のWebアーカイブスと、そこで開発されたWebリンク解析アルゴリズム *Companion* [2] を用いて、ジェンダー研究関連Webコミュニティの抽出と分析を行った。分析にあたっては、ジェンダーが社会的・文化的な所産であるという特質に留意して、ジェンダー概念とそのコミュニティが時代の流れと共に通時的にどのように変遷しているかを、ジェンダー研究の専門家に細かく分析してもらい、新たな知見の発見に努めた。このアプローチは、これまでややもするとアルゴリズムやシステムの開発に重点が置かれてきたデータベース研究の分野に、コンテンツ指向の研究の重要性をアピールして一石を投じることができたのではないかと考えた。もちろん、Webリンク解析アルゴリズムは *Companion* が依ってたつHITS法以外にもさまざまな手法が提案されているが、いずれも、アルゴリズムを特定の応用分野に深く適用した場合に、使用に耐えうるものであるか否かは検証されていない。本研究では、ジェンダー関連Webコミュニティの分析に *Companion* を選択したことで、いくつもの有益な分析が行えたと評価する一方、Webコミュニティの発展過程を読むにあたっては、使用するマイニングツールのビューアとしての特性を熟知しておくことも必要であることを明らかにしている。

2. ジェンダー関連 Web コミュニティの分析 環境と分析目的

2.1 Web アーカイブスと Web マイニングツール *Companion*

・Web アーカイブス

Webコミュニティを分析するには、探索 ロボットを用いてWebページを収集しておく必要がある。そのために、東京大学生産技術研究所喜連川研究室では1999年以来、年に一度のペースで“jp”ドメインのWebアーカイブスを過去5年にわたり収集している。本研究ではそれを使用している。図1にこれらアーカイブスの統計値を要約する[3]。若干の補足をすれば、総ページ数にはjpドメインのWebページからリンクされたjpドメイン以外のページも含む。総シード数は、

*Companion*を稼働させる場合にシードページが必要になるが、その数である。ユーザがコミュニティ抽出にあたり、人為的に与えることもできるが、ここでは“入りリンク”

(in-link) が3本以上のページをすべてシードページとしている。総コミュニティ数はこのシードページのもとに同定されたコミュニティの総数である。

年	収集期間	総ページ数	総URL数	総リンク数	総シード数	総コミュニティ数
1999	7~8月	17M	34M	120M	657K	79K
2000	6~8月	17M	32M	112M	737K	88K
2001	10月初旬	40M	76M	331M	1404K	156K
2002	2月初旬	45M	84M	375M	1511K	170K
2003	2月初旬	34M	82M	338M	1487K	181K

単位 M: Million K: Kilo

図1. 使用したWebアーカイブの統計値

Fig.1 Statistics of Web Archives Used

・Webマイニングツール*Companion*-

Webコミュニティを発見し分析するマイニングツールには、喜連川研究室の豊田正史氏らが開発した*Companion*-を使用した。Webコミュニティの共時的(synchronic)分析が可能な「Webコミュニティブラウザ」[4]と、通時的

(diachronic)分析が可能な「コミュニティ発展過程ビューア」(Community Evolution Viewer, 以下ビューア)[3]の機能を備えている。ビューアでは、「Main History」と

「Detailed History」モードでの表示ができ、両者を使い分けることで、より詳しい分析が可能となる。*Companion*はHITS法[5]に基づいたWebリンク解析アルゴリズムなので、多くのWebサイトからリンクされているオーソリティの存在と、関連する多数のWebサイトへリンクを張るハブの存在がコミュニティ抽出に重要な役割を果たし、シードページ群を与えることにより、ある共通するトピックに関心があるWebページの集合としてWebコミュニティが抽出される。また、抽出されたコミュニティには、コミュニティを構成するWebページのURLに対応しているアンカーテキストから切り出されたキーワード群が出現頻度順にランク付けされて「見出しタグ」として付与される。コミュニティやその識別子の具体例は3章で示す。

2.2 ジェンダー関連 Web コミュニティの分析目的

ジェンダー(gender)とは、社会的・文化的に形成される性別のことを言う(広辞苑)。この概念は1960年ごろ米国で生まれ、男女共同参画社会の実現に向けたさまざまな活動の原動力となっている。ジェンダーは、学問としてとらえてみると、これは大変に若い学問であると同時に、社会的・文化的な特性を有するが故に、短時間のうちにさまざまな事象と関連して急速に変貌しつつ発展しているという意味で、その姿と進化過程をできるだけ客観的に捉えることができるならば、それは学問上大変興味深いことである。幸いに、現代はインターネットの時代であり、現在多数のジェンダー関連サイトをWeb上に見ることができる。

さて、ジェンダー研究がどのような姿で変遷を遂げてきたのかを明らかにするには、次の二つの視点からジェンダー関連Webコミュニティを分析することが必要である：

- (1) Webコミュニティの「共時的」分析、つまりある時点におけるコミュニティの相互関係のスナップショット分析
- (2) 「通時的」分析、つまり時間の流れと共に Web コミュニティがどのように変遷してきたかの分析

項目(1)に関しては、すでに筆者らによりその基礎的考察が行われその一部を報告してきた経緯がある[6]ので、本論文では項目(2)を重点的に研究する。

加えて、分析にあたっては、組織や個人が活動している“実世界(real world)”でのコミュニティと、そこでの活動が“サイバースペース”へ情報発信されて、その結果形成される Web コミュニティの関係に注目する必要がある。したがって、ジェンダーに限らず、コミュニティの(共時的・)通時的分析にあたっては、次の視点が必須となる。

- 実世界コミュニティではよく知られている現象が、Web コミュニティに反映されているか(ビューアの実世界認識能力)
- Web コミュニティでの現象を分析することにより、今まで明確に意識されていなかった実世界コミュニティでの現象を何か発見できるか(ビューアからの発見)
- どう説明してよいか分からない Web コミュニティでの現象はないか、つまりビューアの分析機能あるいは「読み」に問題点はないか(解読法)

図2に実世界のコミュニティ、サイバースペースのコミュニティ、および分析者による認識(feed-forwardの関係)と発見(feedbackの関係)の関係を示す。

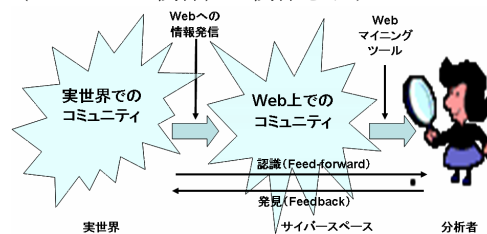


図2 Web マイニングの概念

Fig. 2 Web Mining Concept

3.ジェンダー関連 Web コミュニティの通時的分析

3.1 ビューアの実世界認識能力

上述の視点(a)に示した問題を実世界のジェンダーコミュニティでよく知られている現状に着目して検証する。例として、各地の「女性センター」関連(2003/02のコミュニティ識別子 35392, 具体的には{女性, センター, 男女, 参画, ...}というキーワード群が見出しタグとして付与されている)のコミュニティを取り上げ、そのコミュニティの発展の系譜と「ジェンダー」という語彙との関係から、実世界の現象をWeb コミュニティはどのように反応したかを見た。

図3は、基本となるMain Historyとしての表示法のもと、2003年2月のWebアーカイブを基点に、ビューアにキーワード「ジェンダー」を日本語で入れることで得られる、過去5年にわたる女性センター関連コミュニティの発展過程とそのページ群を示している。ここで、コミュニティの発展過程は、(2003年2月を起点に)共通するURLの数を基軸にして、年次ごと段々に過去(左方向)にさかのぼって、脈絡を持たせていくことで得られている。その際、移動したコミュニティの数が線の太さで表されている。「ジェンダー」をキーワードとしてコミュニティを求めた場合には、2003年2月時点で21個のコミュニティの数が出現する。そして、キーワードである「ジェンダー」の出現頻度が高いコミュニティほど表示ウィンドウの上部にコミュニティが描かれる。



図3 「女性センター」コミュニティの発展過程
Fig. 3 Progress of "Women's Center" Community

この Web コミュニティが発展した様子は、「女性センター」という施設が、各都道府県ばかりでなく各市町村のレベルまで、相互視察によって横並び状態で次々と作られていき、連携を図ってきた現実とぴったりと重なる。各自治体では、十数年前から女性のための拠点作りがはじめられたが、1999年6月に施行された男女共同参画社会基本法の影響で、複数の部局にまたがっていた女性施策部局が、実務部局としての「女性センター」に収斂される動きが加速度的に進んだ。多くは次々と「〇〇女性総合センター」などと改称し、男女共同参画社会を推進するに積極的であることをアピールすべく、その旨をホームページの公開によって示そうとした。2001年の段階で「男女平等推進に関連する施設」の一覧表[7]でホームページの有無を確認すると「〇〇婦人会館」には URL の表示がない場合が多く、「〇〇女性センター」にはほとんど表示がある。また、各施設は独自性を打ち出す名目で、カタカナやひらがなによる愛称をつけた。キーワードの集合体には地域名とともにそれらが見出せる。また、キーワードからは2000年のところで一度、「婦人」を含むが、その後は、ほとんど「女性」に統一されて表記されてくることが確認できた。「女性」を含んだ施設名は、男女共同参画という点からすると、男性が使いにくいという意見もあったが、男女平等が実現していない段階で、女性センターという名称でもいいということで、各地で進められていた。2000年の時点で県立の女性センター建設中が10県、センターを持たない県は未だ17県もあった[8]。ところが、2003年ごろから、「女性センター」が「男女共同参画センター」に改称する動きが出てきている。この傾向については、男女という名をつけたとたん「女性地位向上プログラム」の役割は、もう必要ないとして、財政難から予算を削減する動きや、女性教育事業費を削られるなどの現実に関連していくと懸念している研究者もいる[9]。共通する URL でコミュニティを繋ぐこのビューアは、ホームページのタイトル表記の変化の認知に威力を発揮するのである。

ところで、「女性センター」のコミュニティは、「ジェンダー」というキーワードを投げることで抽出されたコミュニティ21個の中で、最下位に配置されていた。これはコミュニティの見出しタグ、つまりアンカーリンクの文字列に「ジェンダー」の出現頻度が少なかったためである。この位置関係は現実の世界で「男女共同参画社会基本法」の成立過程で「ジェンダー」という概念を基盤とし、構想の出発点としながらも、最終的に「ジェンダー」というカタカナ語は削られたことと直結している[10]。男女共同参画社会基本法の審議段階で基本理念となっていた「性別(ジェンダー)に縛られず」という文面は「性別にかかわらず」と表記され、男女という性別二分法が固定化していく潜在性を残したのである。その影響

はジェンダー関連の Web コミュニティの中でも、とくに設立母体が公的機関のサイトの場合に強く見られるのである。

3.2 ビューアからの発見

視点(b)に示した問題を検証する。「発達過程ビューア」の基本となる「Main History」としての表示のされ方は、横並びに経年の構成 URL の数量的変化が表われることは先に述べた。ほとんどのコミュニティの構成サイト数は年々増加しているが、なかには、数量的に2003年2月時点で規模を縮小している場合がある。たとえば「セクハラ」関連の2つのコミュニティがそれである。今回観察された現象を図4に示すが、これらは過去に遡るコミュニティ展開を見ると、過去4時点のコミュニティは他方のコミュニティの過去と一致している。つまり、サイト数が縮小したかに見えたのは、最新時点において内容的に分裂したためであって、過去では、内容的に未分化であったために、ひとつのコミュニティとしてまとまりを見せていた。これについては、より詳細にサイトの去来を描き出す「Detailed History」モードで見ると、分裂した過去の状況をよりはっきり判断できるが、2003年2月時点には、それまではひとつであった「セクハラ」関係のサイトが、(1)大学関係のキャンパス・セクハラサイト、およびそれらと組織的に同等と見られるサイトをまとまりとして一つのコミュニティとなったコミュニティ、と(2)セクハラ問題への興味関心を求めた個人的サイトをまとまりとした別のコミュニティ、へ分裂したことが見て取れる。(2)は大学関係のキャンパス・セクハラなどの組織的リンク関係から排除されたサイトというべきか、話題に特化したサイトというべきか表現は難しいが、Web コミュニティの分析を通して、実世界でのコミュニティの分裂の「現実」が見えている。これは、Web マイニングツールにより突きつけられた現実であり、発見の範疇で語られるべきものである。

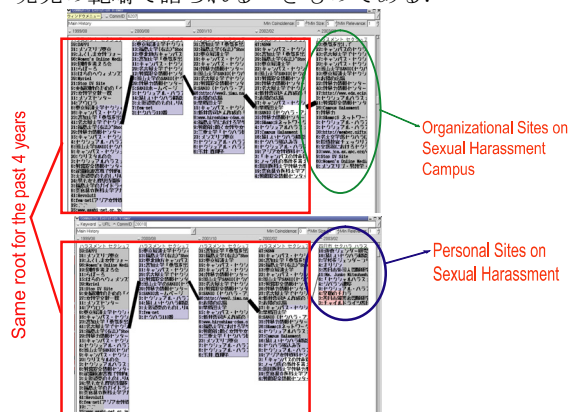


図4 同根コミュニティ系列の分化の発見

Fig. 4 Discovery of Two Different Communities with the Identical Root Community Track

3.3 Web コミュニティの発展過程の解釈法

視点(c)に示した問題を検証する。発達過程ビューアにおいて、結び付けられて表示されているコミュニティ群に付与された見出しタグの文字列は、一般に同じとは限らない。たとえば、図5に表されるように、そこに表示されているコミュニティの発展過程はそれぞれに異なるタグを持つ。詳細に見ると順番や内容に変動が生じていることも確認できる。さらに、注意すべきは、図5に典型的に示されているのだが、横方向に履歴として展開されるコミュニティの見出しタグ中のキーワードに関し、より詳細な一覧を確認しても、どこまで遡っても「ジェンダー」というキーワードが含まれない場

合が生じている点である。なぜこのようなことが起こるのか、これは問題ではないのか。

これはビューアが、共通する URL の数を基軸にして、共通性の高いものをたどって、過去にさかのぼるシステムであるから当然起こるべくして起こったのだという見解もあるが、我々は、かえって、ここに、このビューアを使用したコミュニティ分析の新たな可能性が秘められているのではないかと考えた。たとえば、図 5 の場合、2003 年のアーカイブを始点としたものであり、表示されているのは、SF ファン同好者たちのコミュニティの変遷である。この 2003 年 2 月のアーカイブにおいて「ジェンダー」というキーワードが生じたのであるが、その直接の原因は「ジェンダーSF 研究会」という「既存の男女といった単純な性差に二分化されない、自由な発想」を空想的な世界において探求しようとする会のページが、コミュニティのメンバーに加わったことによる。つまり、このように新たなページが加わり、「Web コミュニティ」の構成する意味合いが新たに付加されていく場合があるのである。よって、キーワードの集合体の変容から、その時々コミュニティの志向性、次の「関心空間」を読み取れる可能性を見出し、その変化を読み取る意味はある。ちなみに、図 5 の過去の履歴というのは、ジェンダー研究者としてはキーワード「ジェンダー」を生み出す土壌を再認識させてくれるものであり、得がたい資料なのである。

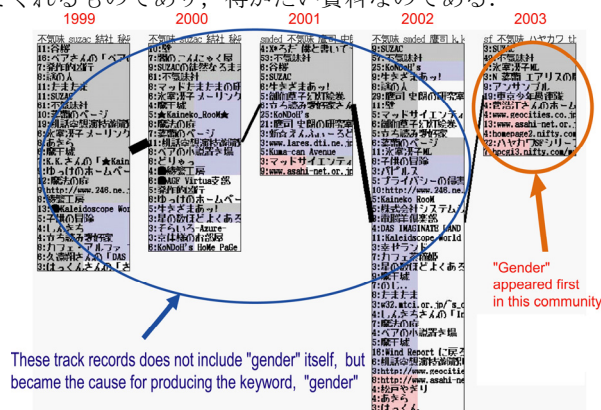


図 5 キーワード「ジェンダー」を含まないジェンダー関連コミュニティの発展過程

Fig. 5 Progress of a Gender-related Community without Having Keyword "Gender"

4. まとめと今後の課題

本研究では、Web コミュニティの分析ツールを用いて、専門家により、その専門分野の Web コミュニティの発展過程をつぶさに分析した。応用ドメインとしては、時間の経過とともに時々刻々とその姿を変えていくジェンダー概念を取り上げた。世界は男女共同参画社会の実現に向けて大きく動き出しているが、ジェンダーはそのキーコンセプトである。分析の結果、その分野の姿を主観的にではなく、客観的に捉えることができ、また、事実の追認だけでなく、これまで明確には認識されていなかった現象をも発見することができたということ、大きな意味があった。加えて、分析ツールには固有の癖（＝特質）があり、それを十分に考慮しないと良い分析結果は得られないことも具体的に明らかにした。ツールが依って立つコミュニティ抽出手法はさまざまであるが、それらの特徴は定かではない現状もあり、ツールが整備されれば、今後さらに分析を試みる価値が十分にある。今回使用した *Companion* のようにハブとオーソリティを基本に

にしたリンク解析によって Web コミュニティを抽出するアルゴリズムが、Web の空間構造の「完璧な分類」を本来指向していないことを意識化しておくことは重要なことである。もちろん、これによって抽出される Web コミュニティは、自己評価も加わった、他者からの認知のされ方を知るのに都合がよいものであり、現実社会においての利用価値は、現時点において既にある。

【謝辞】

本研究を遂行するにあたり多大なご協力とご助言をいただいた東京大学生産技術研究所教授喜連川優氏、同特任助教授豊田正史氏、お茶の水女子大学ジェンダー研究センター教授館かおる氏に感謝する。本研究は一部、日本学術振興会科学研究費補助金基盤研究(B)(2)「Web コミュニティの動的解析手法を用いたジェンダー研究ポータルサイトの構築」(課題番号 15300031)の補助を受けて行われた。

【文献】

- [1] 館かおる, “ジェンダー概念の検討,” ジェンダー研究, 1 号 (通巻 18 号), pp.81-95, お茶の水女子大学ジェンダー研究センター, March 1998.
- [2] M. Toyoda and M. Kitsuregawa, “Creating a Web Community Chart for Navigating Related Communities,” Proc. Hypertext 2001, pp.103-112, 2001.
- [3] 豊田正史, 喜連川優, “日本におけるウェブコミュニティの発展過程,” 日本データベース学会 Letters, Vol.2, No.1, pp.35-38, May 2003.
- [4] 福地健太郎, 豊田正史, 喜連川優, “Web Community Browser: 大規模 Web コミュニティチャートの可視化,” 第13回データ工学ワークショップ(DEWS2002)論文集, March 2002.
- [5] J. M. Kleinberg, “Authoritative Sources in a Hyper-linked Environment,” Proc. 9th ACM-SIAM Symposium on Discrete Algorithms, 1998.
- [6] 増永良文, 小山直子, “ジェンダー関連 Web サイトのコミュニティ分析とポータルサイト構築—Web コミュニティの関連性から見たグローバル化—,” 「グローバル化とジェンダー規範」, pp.101-122, お茶の水女子大学, 2002.
- [7] 関哲夫 (編), “資料集 男女共同参画社会,” 629p., ミネルヴァ書房, Sept. 2001.
- [8] 大沢真理 (編集代表), “21世紀の女性政策と男女共同参画社会基本法,” 303p., ぎょうせい, May 2000.
- [9] 竹村和子 (編), ““ポスト”・フェミニズム,” 182p., 作品社, August 2003.
- [10] 内閣府男女共同参画局(編), “逐条解説 男女共同参画社会基本法,” 172p., ぎょうせい, Feb. 2003.

小山 直子 Naoko OYAMA

1987 お茶の水女子大学大学院家政学研究科修士課程修了。服飾史が専門。現在、お茶の水女子大学 COE 客員研究員としてジェンダー研究に取り組んでいる。日本データベース学会、日本家政学会、日本感性工学会、日本風俗史学会等会員。

増永 良文 Yoshifumi MASUNAGA

お茶の水女子大学理学部情報科学科教授。1970 東北大学大学院工学研究科博士課程修了、工学博士。データベースシステムとその高度応用の研究・開発に従事。現在、日本データベース学会会長。著書に「リレーショナルデータベース入門 [新訂版]」(サイエンス社)など。