

2014年度 卒業論文

山田正雄ゼミナール

グーグルとビッグデータ  
～グーグルの企業戦略と今後の展望～

日本大学法学部 公共政策学科 4年

学籍番号：1150103

三好広和

## はじめに

---

この論文をご覧になっている方の中で、「グーグル」という名前を一度も耳にしたことがない方はおそらくいないのではないだろうか。検索、メール、スマートフォン用 OS など、多岐にわたってサービス展開を行うこの米 ICT 企業は、現在ウェブ上でインターネットを利用するにあたって、あらゆるユーザーに深く浸透してきている。米統計会社のアレクサ<sup>1</sup>の調査によると、2014年12月時点で、グーグルは世界で最もアクセス数が多いサイトとされている。<sup>2</sup>

地球規模で ICT が浸透している現状は、このグーグルの知名度向上の理由のひとつである。2005年には全世界で10.2億人だった世界のインターネットユーザーは増加を続け、2014年時点で29.2億人に達しており、米国通信機器ベンダーCiscoによると、今後もインターネット人口は増加し、2025年には55億人に達すると見込まれている<sup>3</sup>。

それでは、グーグルはウェブにおいて誕生当時から多くのユーザーに利用されていたのだろうか。元来から ICT の覇者といえば、業務用・家庭用にコンピューターを普及させた米マイクロソフトが挙げられるだろう。現状でも日本国内の2014年上半期ソフトウェア市場においては、同社が首位の売り上げを誇っている<sup>4</sup>。また、検索サービスやフリーメールではヤフーという先駆者が存在したし、現在普及が拡大しているスマートフォン用の OS(オペレーティング・システム)では米アップル社の「iOS(アイオーエス)」が先駆けて市場で人気を博した。

グーグルとは、こうした先駆者たちとどのように渡り合い、現在の地位を築いたのか。その理由を私は「ビッグデータの利活用」という点にあると考える。本論文ではこの点からグーグルの特徴や強み、課題を検討していく。

また、グーグルの企業理念(ミッション)は「(Google の使命は)世界中の情報を整理し、世界中の人々がアクセスできて使えるようにすること」<sup>5</sup>である。この理念を実現するための重要な手段がこのビッグデータであると考えられる。そして、この「世界中の情報」とは、これからグーグルが新しいサービスを展開していくにつれ、どんどん多様化してくるであろう。こうした未来に、グーグルはいかなるサービス展開をしていくのかを展望として明らかにしていきたい。

---

<sup>1</sup> 世界中に存在する全てのドメインのアクセス統計を行っている、米 Amazon の子会社

<sup>2</sup> <http://www.alexa.com/topsites>

<sup>3</sup> <http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h26/html/nc111000.html>

<sup>4</sup> <http://news.mynavi.jp/news/2014/12/17/203/>

<sup>5</sup> <https://www.google.co.jp/intl/ja/about/>

# 目次

---

## はじめに

### 1 グーグルとは

- 1.1 企業概要・概歴
- 1.2 企業理念 ～グーグルの目的とは～
- 1.3 特徴 ～なぜシェアを握れたのか～
  - 1.3.1 「フリー」とは
  - 1.3.2 「オープンソース」とは
- 1.4 ビジネスモデル
- 1.5 サービス展開
  - 1.5.1 検索
  - 1.5.2 Gメール
  - 1.5.3 グーグルマップ
  - 1.5.4 アンドロイド

### 2. ビッグデータとは

- 2.1 ビッグデータとは
- 2.2 ビッグデータの登場背景

### 3 グーグルのビッグデータとは

- 3.1 グーグルのビッグデータを支える技術・思想
  - 3.1.1 グーグル・ファイル・システム
  - 3.1.2 マップリデュース
  - 3.1.3 ハドゥープ
  - 3.1.4 データマイニング
  - 3.1.5 データセンター
- 3.2 グーグルのビッグデータの構成要素
- 3.3 グーグルのセンサーデータ

### 4 ビッグデータの問題点

- 4.1 ビッグデータの商業利用によるプライバシー侵害
  - 4.1.1 DPI
  - 4.1.2 ビッグデータの商業利用に対する法制度
- 4.2 ビッグデータの利活用プロセスごとで起こり得るプライバシー侵害
  - 4.2.1 収集・管理のプロセス
  - 4.2.2 分析・活用のプロセス
- 4.3 グレーゾーン：個人情報の第三者提供

- 4.4 その他の問題点
  - 4.4.1 監視への反発
  - 4.4.2 忘れられる権利
  - 4.4.3 データ収集への嫌悪感
  - 4.4.4 技術革新かプライバシーか

## 5 展望

- 5.1 今後のサービス展開
  - 5.1.1 グーグル・カー
  - 5.1.2 グーグル広告
  - 5.1.3 遺伝子情報、人工知能
- 5.2 データの利活用を巡る対応

## 結びに代えて

### 参考文献

### 参考 URL

# 1 グーグルとは

---

## 1.1 企業概要・概歴

グーグルは、1998年のアメリカで誕生した、ICT企業である。創設者はラリー・ペイジとサーゲイ・ブリンの両名で、「グーグル」という検索エンジンを開発した1995年当時は二人ともまだ大学生であった。この名前は1の後にゼロが100個並んだ値を表す「googol（ゴーゴル）」という数学用語をもじったものである。

グーグルの歴史を発表してきたサービスでまとめると以下のようになる。ちなみに、グーグルはたくさんのサービスを発表、運用してきており、その中には提供開始されたものの普及することなく提供中止したものも存在する。そのため今回は概歴という形で、主要なサービスをまとめるに留める。

2000年にはセルフサービスでオンライン広告キャンペーンを作成できる「アドワーズ」を発表した。今日ではディスプレイ、モバイル、動画の広告を展開している。

2004年には無料メールサービス「Gメール」を提供開始した。巨大な保存容量や検索機能などが好評を博し、現在世界で5億人のユーザーを持っている。また、同年に米国において株式公開を果たした。

2005年には「グーグルマップ(Google Map)」と「グーグルアース(Google Earth)」のサービスを開始した。現在のグーグルマップにはリアルタイムの渋滞状況や、交通機関の乗換案内、道路や道路付近の街並みが詳細に画像表示可能な機能の「ストリートビュー(Street View)」などの機能があり、グーグルアースでは海や月面も閲覧可能になっている。

2006年にはオンライン動画共有サイト「ユーチューブ(YouTube)」を買収した。

2007年には携帯端末のためのオープンプラットフォーム「アンドロイド(Android)」を発表した。

2008年にはウェブブラウザ「グーグルクロム(Google Chrome)」を発表した。

2011年にはソーシャルネットワーキングサービス「グーグルプラス(Google+)」を発表した。

2014年にはウェアラブル端末「グラス(Glass)」を発表し、欧米で一般発売が開始されている。

現在はカリフォルニア州に本社を置き、その建物は「グーグルプレックス(Google plex)」という愛称で親しまれている。この愛称にも語源があり、1の1グーゴル乗という数字「グーゴルプレックス」がそれである。こうしたところからも、企業として途方もなく膨大なデータを扱うという気概が窺える。

## 1.2 企業理念 ～グーグルの目的とは～

前述の通り、グーグルの企業理念とは「(Googleの使命は)世界中の情報を整理し、世界中の人々がアクセスできて使えるようにすること」である。この企業理念からは、グーグ

ルは単なる ICT 企業としての目標(=利益追求や社会貢献)のみならず、世界中に散逸した「情報」を秩序立てようとしていることが窺える。

また、企業理念のほかに「Google が掲げる 10 の事実」という名称で行動理念のようなものを掲げている。以下、それらを挙げグーグルがどのような事業を展開しようとしているのかを明らかにしていく。

- ①「ユーザーに焦点を絞れば、他のものはみな後からついてくる。」
- ②「1 つのことをとことん極めてうまくやるのが一番。」
- ③「遅いより速いほうがいい。」
- ④「ウェブ上の民主主義は機能します。」
- ⑤「情報を探したくなるのはパソコンの前にいるときだけではない。」
- ⑥「悪事を働かなくてもお金は稼げる」
- ⑦「世の中にはまだまだ情報があふれている。」
- ⑧「情報のニーズはすべての国境を越える。」
- ⑨「スーツがなくても真剣に仕事はできる。」
- ⑩「『すばらしい』では足りない。」

(引用 : Google が掲げる 10 の事実)<sup>6</sup>

これらの行動理念から企業理念と関連するとして特に注目したいのが、⑦である。この理念の説明には「Google の研究者たちは、今後も世界中のあらゆる情報を検索ユーザーに提供するために開発を続けていきます。」とあるのだが、このようにグーグルの扱う「情報」とはまだまだ完全なものではないことが窺える。

私は、この企業理念と行動理念にある「情報」という言葉に非常に興味を抱いた。なぜなら、一口に情報といってもそれが「どのような」情報なのかによって、グーグルの目指すゴールは変容してくるからである。もしグーグルが検索サービスのみを提供する企業だとするならば、ここでの「情報」とはウェブ上のデジタルデータのみがそれにあたる。しかし現状ではグーグルは非常に多岐にわたってサービスを展開し、その中にはデジタルデータだけでなく地図による現実世界の地理情報やサービスごとのユーザー利用履歴など多様なデータ含まれてくる。こうした多種多様で大量のデータは現在、「ビッグデータ」と呼ばれている。

こうした理念を実現すべく、グーグルはその存在感を日に日に高めているが、なぜグーグルは現状のシェアを握れたのだろうか。次節ではその理由を、二つのグーグルの特徴を明らかにしながら論じていく。

---

<sup>6</sup> <https://www.google.co.jp/intl/ja/about/company/philosophy/>

### 1.3 特徴 ～なぜシェアを握れたのか～

本章では、グーグルはなぜ世界中で存在感を示すことができるようになったのかについて考察を進める。私はその理由を「フリー」と「オープンソース」という二点にあると考える。

#### 1.3.1 「フリー」とは

まず、「フリー」について明らかにする。米 3D ロボティクス社 CEO のクリス・アンダーソン氏が米雑誌ワイヤードの編集長であった 2009 年に出版した著書<sup>7</sup>においてフリーという概念の重要性を世に知らしめた。アンダーソン氏の主張をまとめると、今日の競争市場では、価格は限界費用まで落ちていて、テクノロジー(情報処理能力、記憶容量、通信帯域幅)の限界費用は年々ゼロに近づいていく。そしてアイデアから作られるデジタル商材の開発コストも低下している。それゆえに低い限界費用で複製や伝達出来る情報は無料に近づき、限界費用の高い情報は効果になる傾向があると述べている。こうした状況を「フリー」または「フリー化」という。

こうした「フリー」をグーグルはいち早く追い風にして、ICT 業界でシェアを獲得したといえる。ここではその例として 2004 年から提供が開始された無料メールサービス「G メール」を取り上げる。

G メール登場以前にもヤフーやマイクロソフトが提供する無料メールサービスは存在したが、2004 年のエイプリルフールに発表された本サービスの機能、特にメールの保存容量には世界が驚かされ、悪い冗談とも受け取られた。具体的には、当時最大のユーザー数を抱えていたヤフーの保存容量でも、無料版では 10 メガバイト、有料版の最大でも 100 メガバイトであった。そして、当時のユーザーのほとんどは無料版を利用していた。

こうした状況において G メールは、なんと無料版で 1 ギガバイトの容量を利用可能にするという発表をした。ヤフーもそれに対抗し、2004 年終わりまでには同様に 1 ギガバイトまでを無料提供した。Gメールの登場によって脅かされたヤフーメールであったが、素早い対応や既存ユーザーの愛着などから、ユーザー数の激減という最も恐れていた事態は免れることが出来た。その一方で年々ユーザー数で G メールに追い上げられ、正確な数字が公表されている二〇一二年の数字では、Gメールは 2 億 8790 万、マイクロソフトが運用していたホットメールは 2 億 8620 万、ヤフーメールは 2 億 8170 万のユーザーを獲得している<sup>8</sup>。ちなみに、サービス開始 10 年目にあたる 2014 年 4 月、Gメールのアクティブユーザー数は全世界で 5 億人を超え<sup>9</sup>、無料容量は 15 ギガバイトにまで拡大した。

<sup>7</sup> クリス：アンダーソン(編集：小林弘人) 『フリー～(無料)からお金を生み出す新戦略』 2009 年

<sup>8</sup> <http://gigazine.net/news/20121101-gmail-beats-hotmail/>  
デスクトップ PC からのアクセスのみで、モバイル端末からのアクセスは除かれている。

<sup>9</sup> <http://www.cnn.co.jp/tech/35045970.html>

現在では、メールや地図だけでなく、文書や図表作成サービス、クラウドストレージなど多様で利便性の高いサービスを無料で利用することが出来る。

このように、グーグルは優れたサービスを「無料で」提供することでシェアを獲得していくことが分かる。こうしたビジネスモデルにより、創業10年でグーグルは200億ドルの売り上げを誇る大企業へ成長した。利益も2008年には40億ドルを超えた。

フリーという戦略は、グーグルにとってどのようなメリットがあるのかをまとめると次のようになる。グーグルの便利で有用なサービスを無料で多岐にわたり継続的に提供することで、グーグルならびにウェブを利用するユーザーが増加する。それはすなわちグーグルの市場であるウェブの拡大に繋がり、それに乗じてさらにグーグルのサービスは充実し、ユーザーを呼び込むことが出来る。グーグルのサービスは様々な場面でウェブの利用を促進するものとみなすことが出来るため、さらにグーグルの事業を補完、促進する材料となっているのである。

### 1.3.2 「オープンソース」とは

次なる特徴とは、「オープンソース」であることである。オープンソースとは、だれでも開発に参加でき、だれでも無料で使えるというソフトウェア開発の仕組みのことである。

外部のソフトウェア技術者が開発をすることで、より新たな側面からソフトウェアの利便性が向上するというメリットがある。ここではマイクロソフト社との比較から、オープンソースの特徴を明らかにし、グーグルのオープンソースなサービスについて論じる。

従来からIT産業には、多くのユーザーが一つの製品を使いだすと事実上他の製品は参入できなくなるという寡占構造がある。マイクロソフトがOSにおいてこうした状況を招いた原因の一つは、ソースコード<sup>10</sup>を公開しなかったためである。ほとんどのユーザーがマイクロソフトのOSを使いコンピューターを利用していた状況があったため、そのOSに対応したソフトウェアを作ることを迫られたマイクロソフト以外のエンジニアたちは、ソースコードが分からないのでどうしてもマイクロソフトを超える製品を作り出すことが難しくなった。こうしてますますマイクロソフト製品の寡占が進むという構造があった。<sup>11</sup>

こうした状況においてグーグルは、同じ市場制覇といってもマイクロソフトとは異なった戦略を取っていた。サービスは「フリー」で「オープン」なため、それらのサービスを元にさらに拡張したサービスを社外のエンジニアでも開発できるようになっている。一部でもグーグルオリジナルの要素が用いられていれば、それはグーグルの評判や信頼を高めることにも繋がるという戦略である。

また、この戦略において重要な考えが、「API(アプリケーション・プログラミング・イン

<sup>10</sup> コンピュータプログラムやソフトウェアの元になる文字列のこと。

<sup>11</sup> 参考：『現代思想 一月号 第三九巻第一号』 青土社 2011年 43頁

ターフェイス)」である。APIとは「アプリケーションの開発者が、他のハードウェアやソフトウェアの提供している機能を利用するためのプログラム上の手続きを定めた規約の集合」(総務省)である。

オープンソース性を持つグーグルのサービス例として、ここではグーグルマップを取り上げる。

グーグルは、グーグルマップというウェブ上の無料地図サービスのAPIを公開して、ほかのウェブサイトでもグーグルマップのデータが自由に使えるようにした。他のサイトや企業はグーグルのAPIを利用し自社独自のサービスを開発することができる一方で、グーグルマップはウェブ上のスタンダードになり、さらにグーグル以外のサイトからもデータを収集可能になったのである。

APIが必要な理由は、世の中のデータは広く連携されてこそ、そこから得られる価値が最大化するためである。データを収集し蓄積することに長けた事業者と、それを活用し新しい情報の抽出と付加価値の提供に長けた事業者が登場することとなる。それらを結びつけるのが、APIである。グーグルは、自社でのデータ収集・活用のみならず、他社のサービスにおいてもグーグルのサービスのAPIを利用してもらうことで多岐に渡ってデータを収集することが出来る。

#### 1.4 ビジネスモデル

ここで疑問になるのが、「グーグルはタダでコピー可能な製品ばかりでどうして赤字にならないのだろうか」ということである。企業活動には資金が不可欠である。資金を生み出すには収益を生み出さなくてはならないが、グーグルの収益源とは何か。その答えは、広告収入である。それらは「アドワーズ」や「アドセンス」と呼ばれる広告サービスである。

アドワーズとは、グーグル検索の結果画面に表示される広告であり、検索ワードに関係した広告である。広告主が広告費を支払い利用する。

アドセンスとは、提携した個人のウェブサイトやブログなどに表示される広告である。

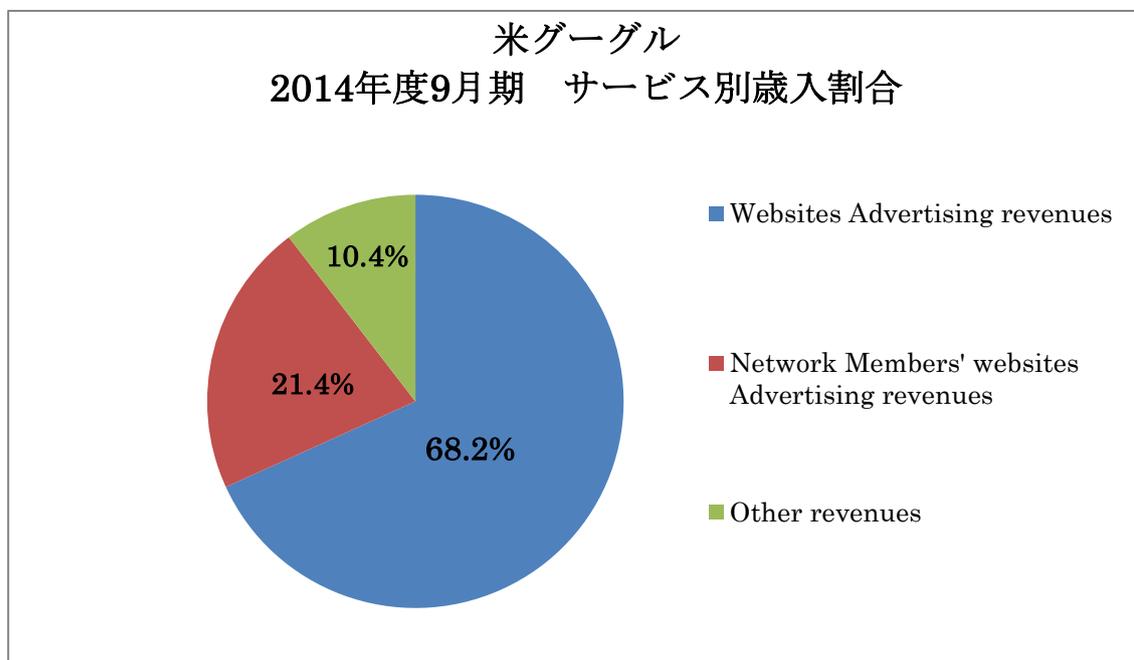
従来の広告とは、たとえば雑誌や新聞の場合、記事との関連性を強くしすぎると読者の反発を受けるため関連性を薄くしたり、一つの号に同じ商品の広告を複数掲載しなかったり、ページ全体のテーマやデザインに沿ったものをバランスよく配置するなどの工夫を行っていた。つまり、記事と広告の関連性が高すぎると、むしろ読者から反発を受けるため、あえて本文内容と広告との関連性を薄めている。

しかしグーグルの広告はこれら従来の広告とは異なることが分かる。グーグルの広告では、検索結果やユーザーの利用履歴に合わせた広告が表示される。これについて不快感を持つユーザーもいる一方で、こうした広告を便利な機能と捉えて積極的に利用しているユーザーもいる。従来のオフライン広告は、その商品に関心がありそうな僅かなユーザーに届けるために、関心がない残りの大多数のユーザーにも広告を届けざるを得ないものであ

ったので、これと比べるといかに効率的かが分かる。

こうした広告収入も、結局グーグルのサービスを無料で幅広く提供することで実現している。広告の反響率が非常に僅かであっても、ウェブの世界では「ロングテール」という考えがある。通販大手アマゾンの事業がこれの例として知られているが、たとえ一年に一人しかその商品を購入しなくても、こうした商品の販売が世界中のウェブで行われれば、塵も積もれば山となる、たとえヒット商品でなくとも収益を生み出すことが出来るという考えである。つまり、ユーザー数を多くなり利用回数が増えるほど、個々の利用は僅かであっても収益を上げることが可能になるという構造を持っている。

ここでグーグルの財務指標を参照し、グーグルの収益におけるサービスごとの割合を確認し、広告収入がいかに多いのかを明らかにしたい。



(図1：筆者作成<sup>12)</sup>)

図1から、グーグルの歳入(revenues)の89.6%は広告(Advertising revenues)であることが分かる。ちなみに、Websites Advertising とは”Paid clicks on Google websites, which include clicks related to ads served on Google owned and operated properties across different geographies and form factors, including search, YouTube engagement ads like TrueView, and other owned and operated properties like Maps and Finance” とあるため、グーグルによる検索やユーチューブやマップなどのサービスからのクリックによる収入とみられる。また、Network Member’s websites Advertising とは”Paid clicks on Google Network Members' websites, which include clicks related to ads served on non-Google

<sup>12</sup> 参考： [http://investor.google.com/pdf/20140930\\_google\\_10Q.pdf](http://investor.google.com/pdf/20140930_google_10Q.pdf)

properties participating in our AdSense for Search, AdSense for Content, and AdMob businesses”と説明されているため、アドセンスやアドワーズによる広告収入がこれに当たると考えられる。<sup>13</sup>

こうしたビジネスモデルについて、グーグルの元 CEO のエリック・シュミット氏は、CEO 当時のインタビューにおいて「グーグルニュース」<sup>14</sup>というサービスを例に、次のように述べている。

「グーグルニュースに関する初期の調査では、それを利用する人は、そのあとで検索を利用した時に検索連動型広告をクリックする割合が、平均の二倍あるとわかりました。それは素晴らしいことで、グーグルニュースは集客のためのサービスになります。それは世界中に提供しているサービスです。でも、もっと洗練された見方をすると、私たちが与えているのはグーグルニュースではなくてグーグルそのものなのです。全てはユーザーをグーグルに関わらせることにかかっている、それを実現してユーザーを獲得できれば、それは最後に収益化につながり、全体としてうまく回るのです。」

このように、事業を拡大し、様々なモノやサービスを生み出し、グーグルと消費者の結びつきを強くして他のモノやサービスにも適応していき、それらの利用に際して広告で収益を得るとというのが、グーグルという企業のビジネスモデルである。

## 1.5 サービス展開

### 1.5.1 検索

ここからは、グーグルの主要なサービスについて考察していく。ここではまず、グーグルの代名詞というべき検索サービスを扱いたい。

はじめに、グーグル登場以前の検索サービスについてまとめていく。インターネットの性質として、複数のコンピューター間で相互に情報をやりとりすることが前提となる。そしてこれが全世界へと広がったことで、多様な情報がネット上に公開されるようになった。そのため、目的の情報へのリンクを知らないネットユーザーたちは、自らの求める情報を得るためには「検索」を行い最適な情報を探し出す必要があった。

当初の検索エンジンは、1990年代の半ばにインフォシークが始めたサービスが発端とされている。この当時は「ロボット」と呼ばれるプログラムをインターネット上に放ち、世界中のホームページ上のデータを収集する。こうして集められたデータは、内容や重要度をアルゴリズム<sup>15</sup>によって自動判別し、巨大なデータベースに保存される。検索エンジンは、利用者が入力した検索キーワードに適合するホームページのデータを拾い集めて、検索結果として表示するのである。

<sup>13</sup> [http://investor.google.com/pdf/20140930\\_google\\_10Q.pdf](http://investor.google.com/pdf/20140930_google_10Q.pdf) P36

<sup>14</sup> グーグルが提供するニュースサービス。記事自体は他社の記事を掲載しており、表示される記事はアルゴリズムによって自動的に選択されている。

<sup>15</sup> 質問を受け取って結果を返すコンピュータのプロセスや数式のこと。

しかし、こうした初期の検索エンジンは、ホームページの内容やキーワードとの関連性を判断する基準が、ホームページ上に記載されているキーワードの多さであるといったように、アルゴリズムがとても単純だったため、不正にアクセス数を稼ぎたいオンラインカジノ運営業者などに悪用されてしまった。

また、90年代後半になるとネット上の情報が爆発的に増えていったため、ロボットによる情報処理が以前のままではどうも追いつかず取りこぼしが多くなり、データベースは不完全になっていった。こうした状況では、検索エンジンを利用しても本当だったらそこにあるはずの目的の情報に行き着くことが出来なくなってしまった。こうした状況のなかで登場したのがグーグルだった。

グーグルはそれまでの検索エンジンとは大きく異なる、革新的なものとして世界を驚かせた。まず、グーグルは「クラスタリング (結合)」というコンピューターテクノロジーを採用した。それまでの検索エンジンは大型コンピューターを導入していたが、爆発的に増加するホームページには追いつけるような拡張性が無かった。なぜなら、大型コンピューターはいったん導入したら入れ替えるのが難しいからである。これに対しグーグルは、普通の安価なパソコンを数千台も並べ、それらを一台の仮想コンピューターとして使えるようにした。これを「クラスタリング」という。一台一台は性能が低くとも、数多く用意すればスーパーコンピューターに匹敵、あるいはそれを超える性能を発揮できる。このような柔軟なシステム構成によりシステム障害や故障にも対応出来るようになった。一台のコンピューターが壊れてもまた新たなコンピューターを追加導入すれば済むので、大規模なシステム障害を避けられるようになったのである。こうしてグーグルは技術的側面やコスト面からも、爆発的なホームページの増加に対応することが可能になった。

次に革新的だったのは、グーグルが「ページランクテクノロジー」という新しい技術を作り出したことである。このテクノロジーは、検索されたキーワードとの関連性をランク付けする際に、「人気のあるホームページからリンクが張られているページ」を優先するという考えである。グーグル以前の検索エンジンもリンクには注目していたが、こちらは単純に「リンクが多いページは最適なページ」と判断してしまっていたため、たとえ内容のない見せかけのリンクであってもたくさん貼られていれば検索結果の上位に表示してしまっていた。だが、このページランクテクノロジーはリンク元がヤフーやマイクロソフトのMSNなど信頼性が高く人気のあるサイトでなければならない。人気のあるサイト、有名なサイトから多くリンクが貼られたサイトは検索結果の上位に表示されたので、グーグルの検索は的確だと評価されるようになった。

現在のグーグル検索には、「シグナル分析」という考えが用いられている。以下、グーグ

ル検索について公式ページ<sup>16</sup>を参考に以下まとめる。

まず、現在のグーグル検索でも従来型のロボットを用いてウェブページ情報を収集している。グーグルは、「ウェブ クローラ」と呼ばれるロボットソフトウェアを使用して、公開されているウェブページをクロール(情報を取得)している。「グーグルボット(Googlebot)」という代表的なクローラは、ユーザーがウェブ上でコンテンツを閲覧する場合と同じように、ウェブページを見て、ページ上のリンクをたどっていく。その過程で、新しいサイトの情報、既存のサイトの変更点、および無効なリンクを重点的に確認していく。これらのリンクやウェブページに関するデータを Google のサーバーに蓄積していく。この情報の蓄積はインデックス<sup>17</sup>形式から成る単語データベースとなる。このインデックスには、単語と単語が使われていたページに関する情報が含まれており、公開された時期、写真や動画の有無など、ページのさまざまな特性に着目し作成されていく。

検索の流れを整理すると次のようになる。まずクロール時にページの情報を収集した後、インデックスを作成することで検索対象の正確な場所を把握していく。そしてユーザーが検索を行うと、まずはアルゴリズムに従ってインデックスから検索キーワードに対し適切なページを見つけていく。

現在のグーグルのアルゴリズムは、200 を超える独自のシグナルを利用している。シグナルには、ウェブサイトのコンテンツのデータ鮮度や特定のサイトにリンクしている他のウェブサイトの数とそのリンクの権限、ウェブページ上の語句、検索キーワードの同義語、サイト上のコンテンツの質、ウェブページの URL とタイトル、ウェブページ、画像、動画、ニュース記事、パーソナル検索結果などが含まれるとされているが、正確な情報は公開されていない。

グーグル検索はこうした多くのシグナルを混ぜて使うことで、ユーザーの要望に合わせて個人的な検索結果を表示可能になっている。一つの検索された言葉に対して多くのシグナルを利用して検索精度を上げているのである。よって、よりたくさん検索がされればされるほどその言葉に対しての情報が蓄積され、人々が求めている言葉のウェブ情報とリンクされていき、グーグルの検索品質は向上している。

そして前章でも述べたように、精度の高い検索サービスは広告により収益を得るサービスにもなった。検索結果に合った広告をユーザー個別に表示させることが可能になったので、グーグルを通じてウェブページを閲覧する際に、広告提供者側にもユーザー側にもメリットを生むことになった。

このように、ロボットとシグナルを活用して検索結果を向上させ、基本無料のサービス

---

<sup>16</sup> <http://www.google.co.jp/intl/ja/insidesearch/howsearchworks/crawling-indexing.html>

<sup>17</sup> 索引、見出し、添字の意。データベースや検索エンジンでは、格納されたデータをすばやく検索・抽出するための識別データ・索引データなどを指す。

展開を支える収益源でもあるのが、現状のグーグル検索である。ここで注目しておきたいのが、グーグルが検索に「多くのシグナル」を利用しているという点である。グーグルで検索される言葉は毎日数十億と極めて多いので、単語に紐付けられるシグナルはさらに多種多様になる。こうしたデータは昨今では「ビッグデータ」と呼ばれている。グーグルは主幹事業の検索で、ビッグデータを利活用していると考えられる。

### 1.5.2 Gメール

次に、グーグルの特徴をほかのサービスから探っていく。まずは2004年から提供が開始された無料メールサービス「Gメール」を取り上げる。

前章で明らかにしたように無料容量の面で先駆的だったGメールだが、これをデータの観点から捉えてみたい。Gメールでは大量のユーザーのメールを独自のアルゴリズムにより人を介さずに解析し広告を打ち出している。これはつまり、5億人のユーザーのメール会話を解析していることに他ならない。検索サービス同様、Gメールにおいてもグーグルは非常に大量の言語に関するデータを利活用していることが分かる。

### 1.5.3 グーグルマップ

2005年には「グーグルマップ」という地図サービスを開始した。このサービスも無料で、世界中の街の路地裏まで閲覧可能で、有料カーナビゲーションのような機能もある世界規模のサービスである。このサービスは現在たくさんのウェブサイトやアプリケーションで住所や建物の場所を表す場合に利用されている。このことで、このサービスにおいても、グーグルが世界中の地理検索データを手に入れたと考えることが出来る。

この地図の中でユーザーが探しているものがどこにあるかということ、ほかの検索などで集めたデータの中から引きだして、地図のデータと合わせていく。初期のころは住所などで検索していたのが、次第に店の名前や会社名だけで地図上の検索が可能になり、また目的地の周辺にあるレストランやガソリンスタンドなどが地図上や検索ボックス内でサジェストされるようになった。そして、前述のようにAPI戦略においても、このサービスの果たす役割は大きい。ウェブ上だけでなくリアルな現実には店舗などを持つ主体にとって、ウェブ上の正確な地図サービスというものは必要不可欠である。グーグルマップはその正確性やAPIが利用可能であることなどから普及が進み、非常に多くのサイトで利用されている。

### 1.5.4 アンドロイド

さらに、グーグルはPCだけでなく携帯端末にも手を伸ばしている。アンドロイドとは、OS、ユーザーインターフェースなど携帯端末を開発する上で必要不可欠なソフトウェアであり、グーグルはこれをオープンソースかつ無償で提供している。この目的は、携帯広告市場でシェアを獲得するためと考えられる。2008年にはアンドロイドを搭載したスマート

フォンが発売された。

アンドロイドを利用することには携帯端末メーカーにもメリットがあるため、アンドロイド端末は端末のメーカーと OS のメーカーとが一致しない場合が多い。無償であるため、携帯端末開発費用の大半を占めるといわれているソフトウェアにおいて大幅なコスト削減が期待出来るからである。

また、アンドロイドには開発段階からクラウドにデータを保存するという特徴がある。クラウドを用いると、クラウド側の処理能力が高ければ端末側の性能は必ずしも高性能である必要がなくなるため、利用可能なアプリケーションの幅が広がるなどのメリットがある。

その一方でグーグルのメリットはなんだろうか。それはすなわち、幅広いメーカーに自社の OS を普及させることで、多様なユーザーを囲い込み、理念をより実現することが可能になることであろう。アンドロイドの普及を推進することで、従来のマイクロソフトなどのように PC を支配するだけでなく、拡大する携帯端末市場をも手中に収めることが出来る。理念実現のため、グーグルにとってアンドロイドとは非常に大きな役割を果たしているといえる。

その他にも、カレンダー、翻訳、自動入力サービス(IME)、グーグルプラス、無料クラウドサービスである「グーグルドライブ」や動画サービス「ユーチューブ」、多様な利活用が検討され実現されつつあるウェアラブル端末「グラス」、ハンドルがない自動運転自動車などがある。これらに通じている、そして今後通じると予想される特徴とは、多種多様なサービスにおける利活用データの収集と利活用、つまりビッグデータを使ったサービスの最適化である。

ヤフーなども多様なサービスで同様のビッグデータ分析を行っていると考えられるが、グーグルは検索サービスやメールサービスなどの比較でも分かるように、先駆者をも飲み込むその勢いとさらにシェアを伸ばし続ける力があり、それゆえにさらに多様で大量なビッグデータを武器に出来る企業である。

## 2. ビッグデータとは

これまで「ビッグデータ」という言葉を多用してきたが、ここではそのビッグデータについてどのようなものなのか、明らかにしていきたい。

### 2.1 ビッグデータとは

近年「ビッグデータ」の利活用がさまざまな組織で検討、実施されている。平成 25 年度版情報通信白書では、ビッグデータの利活用による経済効果は年間約 7 兆 7700 億円に達するとの試算が出ている。その利活用分野はマーケティングだけでなく、防犯や交通インフラの改善などにも及んでいる。ビッグデータによって、従来より質の高いサービスの提供ができ得ることから注目を浴びているのである。

ビッグデータとは、「既存の一般的な技術では管理するのが困難な大量のデータ群」であると定義したい。管理が困難になる要因は、3V(Volume 量/Variety 多様性/Velocity 発生頻度・更新頻度)である(図 1 参照)。つまり、従来よりも大量かつ種類が豊富なデータで、さらにデータが発生・更新されるタイミングが様々なため、従来のシステムでは蓄積・処理・分析が困難であるものを、ビッグデータと呼ぶ。

3つのV	説明
Volume【量】	従来よりも、 <u>大量のデータ</u> を扱えるようになった。
Variety【多様性】	従来よりも、 <u>多様なデータ</u> を扱えるようになった。
Velocity【頻度】	従来よりも、 <u>頻繁に発生・更新するデータ</u> が扱えるようになった。

(図 2 : 筆者作成)

また、ビッグデータの利活用にはプロセスがある。まず、様々な手法で多様なデータを収集し、蓄積し、目的に応じて加工、分析を行い、実際のサービスに活かしていく。

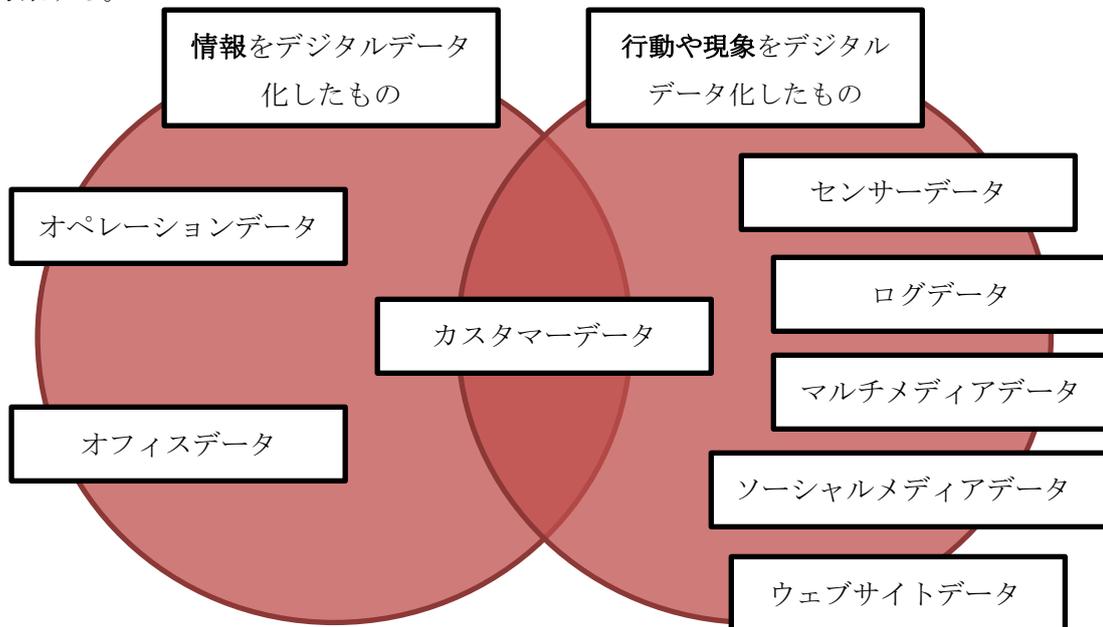
つまり、収集・蓄積→加工・分析→活用というプロセスを経て、ビッグデータは利活用されるのである。

ビッグデータを構成するデータとしては、ソーシャルメディアデータ(ソーシャルメディアにおいて参加者が書き込むプロフィール、コメント等)、マルチメディアデータ(ウェブサイト上の配信等において提供等される音声、動画等)、ウェブサイトデータ(EC サイトやブログ等において蓄積等される購入履歴、ブログエントリー等)、カスタマーデータ(CRM システムにおいて管理等される DM 等販促データ、館員カードデータ等)、センサーデータ(GPS, IC カードや RFID 等において検知等される位置、乗車位置、温度、加速度等)、オフィスデータ(オフィスのパソコン等において作成等されるオフィス文書、Eメール等)、ログデータ(ウェブサーバー等において自動的に生成されるアクセスログ、エラーログ等)、そしてオペレーションデータ(販促管理等の業務システムにおいて生成等される POS データ、取引明細データ等)などが主にある(図 2 参照)。



(図 3 : 平成 24 年度情報通信白書(総務省)より引用)

私はさらにこれらのデータを、「情報」をデジタルデータ化したものと「行動や現象」をデジタルデータ化したものという、どのような対象をデータ化しているのかという視点から考察する。



(図 4 : 筆者作成)

この図で分かることは、ビッグデータの利活用で用いられるデータは、行動や現象をデジタルデータ化したものが多くを占める。つまり、これらを分析することとは、私たちの行動や現実で起こる現象を分析することを意味する。

さらに、米 ICT 企業セールスフォース・ドットコム の代表であるマーク・ベニオフ氏らの著書<sup>18</sup>における定義によると、これらのデータを現実世界における「コンテキスト」ととらえ、それを以下の五つに分類している。

- ①モバイル
- ②ソーシャルメディア
- ③ビッグデータ
- ④センサー
- ⑤位置情報

<sup>18</sup> ロバート・スコープル/シェル・イスラエル(序文：マーク・ベニオフ 訳：滑川海彦・高橋信夫) 『コンテキストの時代 ウェアラブルがもたらす次の 10 年』 2014 年

これらのうち、③はビッグデータそのものに当たり、①、②、④、⑤はデータを収集する手段や、データが生成される土台である。総務省のデータの定義と同様の言葉が定義付けられていることが分かる。

ここでのコンテキストの定義とは、「前後の事情」「背景」を意味している。コンピューターがユーザーの事情や背景を知り、ユーザーが必要とするサービスを的確に判断したり、予測したりできるようになることを指す。つまり、ビッグデータの分析によってこうした未来が実現することを示唆しているのである。

## 2.2 ビッグデータの登場背景

では、こうした以前とは異なるデータをなぜ利活用できるようになったのか、ビッグデータの登場背景を明らかにする。

一つには導入コストの低下がある。データを蓄積する保存容量の価格は年々下落しており、大量のデータを保存して管理することにかかるコストが低下している。(用出典)また、データを処理するコンピューターの処理能力が向上した結果、過去と比べて安価で高性能な設備が導入可能になった。そして、クラウドサービスが普及したことで、設備を自前で用意する必要が軽減された。こうしたことがビッグデータの処理環境を整えた。

また、3Vデータに対応した技術の登場もある。従来の技術では分析することが出来なかった3Vデータの分析に適した、「マッピングリデュース(MapReduce)」、「ハドゥーブ(Hadoop)」等の技術が登場したことも背景である。

つまり、ビッグデータは、従来のデータ活用の延長線上にあるが、3Vという点でより有用性が高まっている。ICTの発展により導入コストが低下し、分析技術が進歩したことで、これまで利用できなかったデータも利活用できるようになった。これまで利用できなかったデータによって人々の行動や現象を分析できるようになり、それらを対象にした利活用が可能になったものがビッグデータである。

## 3 グーグルのビッグデータとは

---

本章では、グーグルはどのようにビッグデータという武器を活用し成長を遂げてきたのかを明らかにし、グーグルのビッグデータというものをさらに探っていく。

### 3.1 グーグルのビッグデータを支える技術・思想

ここでは、グーグルのビッグデータ利活用を支える技術を明らかにしていく。それは、「グーグル・ファイル・システム(Google File System)」、前述したビッグデータの利活用技術である「マッピングリデュース(MapReduce)」と「ハドゥープ(Hadoop)」、そして「データマイニング」である。

#### 3.1.1 グーグル・ファイル・システム

グーグルは元来から、検索のためのウェブのインデックスを膨大な量保存する必要があったため、サーバーのコストを安く抑える方法を編み出した。たいていの企業ではサーバーの故障率をなるべく低く抑えるために堅牢な高品質製品を使用するのに対して、グーグルでは低機能で故障が多いが低価格の製品を集め、故障することを前提として複数のサーバーに処理を分散し、一つが故障しても全体のデータが失われないようにした。このデータ保存の仕組みを「グーグル・ファイル・システム」と呼ぶ。この分散処理の考えはその後グーグルのシステムの基礎となった。

#### 3.1.2 マッピングリデュース

この分散並列処理をより高速に行うためにグーグル内部で開発されたのが、「マッピングリデュース」というソフトウェアだった。これにより、多くのサーバーやデータセンターに分散した大量のデータを高速処理することが可能になり、またサーバーを増設したり交換することが容易になり、急激にデータが増加してもハードウェアの拡張性を増すことが出来るようになった。二〇〇六年にグーグルがこの一連の技術についての論文を公開したことで、グーグル外で関連のオープンソース・ソフトウェアの開発が進んだ。

#### 3.1.3 ハドゥープ

マッピングリデュースは、オープン・ソースソフトウェアを管理する非営利団体「Apache Software(アパッチ・ソフトウェア)」財団傘下のプロジェクト「ハドゥープ」となり、グーグル以外の企業に所属する技術者も数多く参加して開発が進められており、この技術は現在ビッグデータの根幹技術として利用されている。

ハドゥープはデータを処理するためのマッピングリデュースと、データを格納するための「Hadoop Distributed File System(HDFS)」の二つの技術からなっている。マッピングリデュースでは、マスターとなるノード(コンピューター)が、システムにつながったたくさんの他

のノードに課題を分散して割り当て(マップ)、マスターがその答えを集めて課題の回答を求める(リデュース)という手段を実行する。リーダーが采配を振るい、たくさんのコンピューターが皆で手分けして問題を解く仕組みである。また、マップリデュースは、ある程度の基幹にたまったデータを一括処理(パッチ処理)するのに適しているため、グーグルの検索のようなユーザーがキーワードを投入するとこれまでためてきた情報の中から最適な結果を返答するという作業に向いている。HDFS ではデータのコピーを三つ作って異なるノードに分散保存して管理している。これにより一つのノードが壊れてもデータが紛失することではなく、さらにノード処理負荷を分散することもできる。この二つの部分が大きな軸となっているが、この機能以外にも様々なソフトウェアと組み合わせて利用されることもある。

グーグルが開発した技術がビッグデータ分析において主流となっている現状を鑑みると、グーグルとは、自社のサービスがビッグデータそのものであるだけでなく、ビッグデータのコア技術と思想を作り出した存在でもあるといえよう。

### 3.1.4 データマイニング

データマイニングとは、ビッグデータというデータ利活用全般を含む概念の「手法」の一つであると考えられる。

マイニングとは鉱山で鉱物を掘り出すことを意味し、データマイニングとは大量のデータから意味や関連性を掘り出すことである。膨大で多様なデータのなかから、必要なデータや結果を探し出す手法がデータマイニングである。

データマイニングでは、様々な統計分析手法を組み合わせることが多い。その手法をいくつか挙げると、頻繁に同時に発生するデータ同士を関連付ける「相関ルール学習」、類似の傾向を持つデータ同士をグループにまとめてそれらの関係をみる「クラスタ分析」、過去のパターンをもとに新しいデータがどのカテゴリーに属するかを予測する「クラス分類」、多くのデータ点の相関関係の傾向を数式化する「回帰分析」などである。

また、データマイニングと関連深い手法として「機械学習」がある。人工知能の一分野と考えられており、人間が新たな物事を学習するように、過去に蓄積されたデータのパターン解析に基づいてアルゴリズムを発展させ、コンピューターに学習をさせるものである。データマイニングとの違いは、データマイニングは未知の領域を探り出すという指向性を持つのに対し、機械学習は既存のパターンを踏襲するという目的を持っている点である。このほかにも、不定型なテキストや会話をコンピューター・アルゴリズムで分析する「自然言語処理」、ネットワーク内の複数のノード(たとえば、人間関係ネットワークにおいてならば、個々の人間のこと)同士の関係の特徴を探る「ネットワーク分析」、数式モデルで複数の要素同士の最適なポイントを探し出す「最適化」、仮設モデルを動かしてシナリオごとの予測を行う「シミュレーション」、解析結果をグラフ・図・アニメーションなどで表現して理解を助ける「可視化」などの手法がある。

### 3.1.5 データセンター

また、グーグルが持つのはビッグデータの技術だけではない。2006年以降、グーグルは世界中に巨大なデータセンターを次々と建設している。その投資額は年々増加の一途をたどっており、アメリカ国内だけでなくヨーロッパやアジアなど世界中各地に数十億円規模の巨大なデータセンターを建設し、世界中においてデータの受け皿となる土台を着々と構築しつつある。

### 3.2 グーグルのビッグデータの構成要素

ここではグーグルが扱っているビッグデータの構成要素を明らかにしたい。

検索サービスによるデータは、前述の総務省によるビッグデータを構成するデータ分類と照らし合わせると、「ウェブサイトデータ」や「ログデータ」にあたりと考えられる。しかしその一方でグーグルの検索精度を支えるシグナル分析に用いられるシグナルを考慮すると、さらに「マルチメディアデータ」などもグーグルのビッグデータと考えることが出来るだろう。Gメールによるビッグデータも、人語テキストデータを扱っている点を鑑みると、検索サービスとほぼ同様のデータと考えられる。これらと異なるデータでは、グーグルプラスで「ソーシャルメディアデータ」、グーグルマップでは従来のテキストデータだけではなく「地理データ」を扱っている。また、ユーチューブやグーグルクロームなどから収集したユーザーのサービスやウェブの利用履歴も、ユーザーの趣味や思想を含めた大量で多様なデータとなる。

従来のビッグデータの定義では主に顧客にモノやサービスを販売するビジネス上でのデータを指すことが多かったが、グーグルのビッグデータはそれらと比べてさらに多様であることが分かる。

### 3.3 グーグルのセンサーデータ

グーグルのビッグデータをさらに多様で大量にし得るのが、アンドロイドを搭載した携帯端末などから収集されるセンサーデータである。前述のようにアンドロイドとは、携帯端末やウェアラブル端末など、従来のコンピューターだけではなく、様々な端末に搭載されているOSであるが、携帯端末は一般的なPCと異なり、搭載されたセンサーからデータを収集することが出来る。

野村総研の定義では、センサーとは「物理的な事象を、電子データに置き換えるための機械的な仕組み」の総称である。また、ネットワークに接続された自動車や家電などの機器群も、本節においては広義のセンサーである。グーグルのアンドロイドには多様なセンサーが搭載されている。以下それをまとめる。

加速度センサー：加速度を検出するセンサー。シェイク操作などに利用できる。

温度センサー：周囲の温度を検出するセンサー。

重力センサー：重力を検出するセンサー。シェイク操作などに利用できる。

ジャイロセンサー：角速度を検出するセンサー。端末の回転を検知できる。

照度センサー：周囲の照度を検出するセンサー。明るさに応じて画面の輝度を調整するなど利用できる。

線形加速度センサー：単一方向の加速度を検出できるセンサー。

地磁気センサー：磁場を検出するセンサー。コンパスに利用可能。

圧力センサー：圧力を検出するセンサー。気圧をモニタリングすることが出来る。

近接センサー：物体の接近を検知するセンサー。端末に近づくことを検知して画面をロックするなどの利用が可能。

湿度センサー：湿度を検知するセンサー。

回転ベクトルセンサー：回転ベクトルを検知するセンサー。

アンドロイド端末で得られるデータとは、コンピューターとは異なる性質を持つ。GPS位置情報<sup>19</sup>、検索履歴、Gメールなどの操作が自動的にユーザー個人のアカウントと連携するため、個人と深く結びついたデータとなる。さらに端末を利用するシーンはコンピューターのように自室で着座して利用するにとどまらず、外出時、スポーツや娯楽、余暇などでも携帯することが多い。そうした場面での音声、写真、動画などの入力、GPS、前述した各種センサーにより、コンピューターよりも多種多様なデータを収集できるのである。

このことから分かるのは、アンドロイド上でのデータ収集とはすなわち、ユーザーの日常行動により近づいたデータである。たとえば毎日どのルートで通学、通勤しているのかであったり、どのようなタイミングでどのようなアプリケーションを使用しているのかであったり、端末に搭載された様々なセンサーから取得したセンサーデータも含まれる。つまり、グーグルのデータとは業務などで使われるデジタルデータやウェブデータのみならず、よりユーザーの個人的なデータを利活用しているものということを強調しておきたい。

また、今後のサービスとしては今日、「グラス」などの身に着けられる端末、「ウェアラブルデバイス」が注目を集めている。ここでは、ビッグデータ化されてこなかった一般的な生活を営む人々の行動などをデータ化し収集することが可能になる。グラスにもアンドロイドが搭載されることを考慮すれば、さらにアンドロイドから収集されるデータは多様になる。また、次世代のグーグル製アンドロイド端末として注目を集めているグラスにもアンドロイド端末とほぼ同様のセンサーが搭載されている。

### グラスが搭載するセンサーリスト

ジャイロセンサー、補正ジャイロセンサー（物体がどれだけ傾いたか）

加速度センサー、リニア加速度センサー（物体がどれだけ速く動いたか）

---

<sup>19</sup> 人工衛星を用いて経度や緯度、高度といった位置情報を計測するシステム。

磁場（物体の周りにどの程度の磁場があるか）

方位（地球の磁場を計測、物体がどこにあるのか）

回転ベクトル（上記の **android** と同じように、回転ベクトルを計測するもの）

重力（物体にどれだけ重力がかかるか）

光センサー（物体にどれだけ光が当たるか。照度センサーと同じか）

つまり、アンドロイドとはユーザーのコンテキストをより詳しく正確に把握するためであると考えられる。コンテキストを把握し、ユーザーをより深く知るためのインフラとして自社製 OS であるアンドロイドを開発し、そこからビッグデータの収集、分析、利活用がなされ、グーグルという企業活動を支える主要サービスが生み出されるのである。

このように、グーグルのビッグデータとはたった一社のビッグデータであるが、様々なサービスや端末、ユーザーから収集されるデータであることが分かる。従来からビッグデータとして利活用されてきたデータの他に、特に非構造化データのバリエーションが非常に豊富な点が特徴であると私は考える。ウェアラブルデバイスからのデータや今後の自動運転車などからのデータを、従来までの検索やウェブ、サービス利用のデータを、自社のサービス内で連携させることが出来るのは、現時点でグーグル以外に存在しない。こうしたビッグデータはグーグルの武器になる反面、様々で新たな問題を生む原因にもなるだろう。次の章では、こうしたグーグルの影の部分について考察する。

## 4 ビッグデータの問題点

---

いかに機械から生み出されたデータであっても、GPSならば運転者の位置情報、監視カメラの映像データならば個人の姿かたちが記録されている。電気利用料金も、利用料のパターンを見れば在宅中か留守中かや、生活パターンなどが見えてくる。このように、多くのデータが何らかの形で人の属性や行動と関わってくる。ビッグデータといえ、必ずと言っていいほどプライバシー問題が挙げられる。利用者は、個人情報を含むデータを提供し、その対価としてサービスを受けるわけである。これがきちんと守られている限りにおいては、プライバシー問題というよりも個人と事業者間のバランスがきちんととれているかが問題であると考ええる。こうしたバランスは、国や地域、年代や性別、そして個人個人で異なる。バランスが取れているとユーザーが感じている限りにおいては、もちろん何の問題もないが、そのバランスが崩れユーザーが嫌悪感を抱く場合、法令違反が存在する場合、そしてグレーゾーンとなるビッグデータの利活用はどうだろうか。

本章で取り上げるのは、まずビッグデータの利活用における問題点を特にデータの商業利用という点から考察する。そして、個人情報保護法などから法令違反とされる場合をビッグデータの利活用プロセスから考察し、次に法整備が追い付いていないグレーゾーンのビッグデータ利活用、そしてさらに、グーグルのサービスへの嫌悪感などのグーグルのビッグデータサービスにおける影の部分の部分を明らかにし、これらの分析からビッグデータを取り巻く問題を明らかにしていく。

### 4.1 ビッグデータの商業利用によるプライバシー侵害

#### 4.1.1 DPI

現在、ビッグデータの利活用におけるプライバシー侵害として問題視されているのは、ユーザーの様々なウェブ等の利用データを収集され、本人も意識することなかった趣味や傾向の分析をされてしまうという点であるだろう。こうした手法は、ディープ・パケット・インスペクション(DPI)と呼ばれている。DPIの事例としては、ウェブサービス等を利用したユーザーの様々な情報をプロバイダーと呼ばれるネット接続事業者などが収集し、それをもとにDPI業者が利用者ごとに広告を作成し配信するというものがある。

DPIのような個人データ利用による広告は、現在普及が進んでおり、ウェブブラウザなどで閲覧・購入した商品の傾向から、それらに類似した商品が広告として表示されるといった手法は身近になっている。こうした現状に対して、法的側面からはどのような対策がなされているのだろうか。

#### 4.1.2 ビッグデータの商業利用に対する法整備

個人データの取り扱いなど、個人情報保護のための法制度として、日本では個人情報保護法が施行されている。この改正を進める動きのなかで、「何を買ったか」「どこへ行った

か」などの個人データの取り扱いには注目が集まっている。

現行法の個人情報とは、特定の個人を識別できる情報であり、具体的には氏名、生年月日などがこれに該当する。こうした情報に加え、新たに個人情報として法令の保護対象に加えようとしているのが、指紋や顔認識データ、パスポート番号、免許証番号、携帯電話番号などである。こうして現状に合わせ保護範囲を拡充しているが、一方で保護を強めすぎると活発なデータの利活用が妨げられ、革新的なサービスが生まれる機会損失となるリスクもある。そのため改正案では、匿名加工された購買履歴や位置情報の利用においては、本人の同意が無くても活用できるようにしようとしている。

2015年1月時点では、政府の経済成長戦略に伴い、ビッグデータの商業利用を促進していく方向性で議論が進められている。こうした現状において、以下ではビッグデータの利活用の各プロセスにおいて、どのようにプライバシー侵害が起き得るのか、そして現状の法制度におけるグレーゾーンについて論じる。

## 4.2 ビッグデータの利活用プロセスごとで起こり得るプライバシー侵害

### 4.2.1 収集・管理のプロセス

データを収集し管理する段階において、過去にグーグルがプライバシー侵害をしたとして法律違反と認められた事例がある。

イギリスでは、Googleが「Street View」撮影車両によってセキュリティ対策の不十分なWi-Fiネットワークから個人情報データを収集した行為について、同国のデータ保護法に違反していると認定した。企業は、コンプライアンスを徹底し、CSR<sup>20</sup>を全うしなければならないのは言うまでもないが、さらなるプライバシー侵害が起き得るのは次のプロセスであると考えられる。

### 4.2.2 分析・活用のプロセス

ここでの懸念は、分析・活用を行うプロセスにおいて、最もプライバシーの侵害が起こり得るということである。利用者に影響を及ぼさなければ、実害は出ないか、もしくは少ない状態で食い止めることが出来るが、ここで適切にサービスが運用されていないと実害が発生してしまう。次には、グーグルの検索広告に問題があった事例を取り上げる。

グーグルの検索結果に表示される広告の枠に、不正に作成されたショッピングサイトが多数表示されていたことが分かり、グーグルが調査しているという報道がNHKによってなされた。報道によると、家電品の販売を装って不正に作成された多数のサイトが、検索連動型広告として表示されており、こうしたサイトは「家電激安」などというキーワードで

---

<sup>20</sup> Corporate Social Responsibility の略。企業の社会的責任。企業が利益を追求するだけでなく、組織活動が社会へ与える影響に責任をもち、あらゆる企業が利益を追求するだけでなく、組織活動が社会へ与える影響に責任をもち、あらゆるステークホルダー（利害関係者：消費者、投資家等、及び社会全体）からの要求に対して適切な意思決定をすることを指す。

検索すると広告枠に表示され、洗濯機や冷蔵庫などが 8 割引きで買えるなどと掲載されていたが、リンクをクリックして表示されるページは実在する英会話教室や雑貨店、民宿などのホームページを不正に改ざんして作られていたため、広告内容に誤表記があったという。NHK がページを改ざんされた 7 つの事業者と 1 つの団体取材したところ、いずれも家電品の販売は行っていないと回答しており、こうしたグーグルの広告枠への不正なサイトの表示は他にも相次いでいる。グーグルは「指摘された内容について調査している。広告枠への掲載については指針に基づいて審査し、不正が疑われる広告は排除することになっている」と説明しているという。

(参考：NHK ニュース<sup>21</sup>)

この事例では、ショッピングサイトを装い、ユーザーの金銭や個人情報を不正に入手しようとする業者に、グーグル広告が利用されてしまった。こうした状況は、前述した検索エンジンが登場した初期の事例と重なる。グーグル広告の信頼度を高め、現状の対策を徹底する姿勢は常に求められるだろう。

#### 4.3 グレーゾーン：個人情報の第三者提供

ビッグデータの利活用には個人に関わる詳細な情報が含まれる。

##### 個人情報とは

個人情報保護法 第 2 条

この法律において「個人情報」とは、生存する個人に関する情報であつて、当該情報に含まれる氏名、生年月日その他の記述等により特定の個人を識別することができるもの（他の情報と容易に照合することができ、それにより特定の個人を識別することができることとなるものを含む。）をいう。

つまり、ビッグデータの利活用で生成された個人に関する詳細なデータは、個人情報にあたる。こうした情報を、たとえば広告ビジネスなどに活用するために第三者に売買、譲渡する場合、明確な規則が存在する。

第三者提供をする場合には、まずあらかじめユーザーに活用目的を明確に示した後、

- ①それに対して許諾を取ってから提供する→オプトイン方式
- ②許諾なしで提供できるが、利用者からの要請があれば当該データを削除し利活用しない→オプトアウト方式

---

<sup>21</sup> グーグル広告枠に不正サイト多数 NHKニュース web 版 1 月 7 日  
<http://www3.nhk.or.jp/news/html/20150107/k10014514591000.html>

という二つの方式で許諾を得なければならない。

第三者提供の定義を確認すると、グループ会社間で個人データを活用する場合や他の組織間（フランチャイズ組織の本部と加盟店、同業者間等）で個人データを活用する場合である。こうした場合はあらかじめ活用目的を明確に示し、オプトイン方式かオプトアウト方式を設ける必要がある。

その一方で第三者提供にならない場合も存在する。社内で利活用する場合や、同一組織内での情報利用は第三者提供には当たらないため、あらかじめ活用目的を明確に示して同意をとる必要はない。

では問題のグレーゾーンであるが、ここでは事例を挙げて考察していく。

共通ポイント「Tポイント」を運営するカルチュア・コンビニエンス・クラブ（CCC）では、加盟している企業間で、利用者情報を共有してサービスを提供しており、加盟店で集められた購買情報等は、CCCに一元管理されている。加盟店で集められた情報と、氏名や住所などの個人情報統合されると、誰が何を買ったのか、どういった嗜好を持つのか等が特定可能になる。

加盟企業間で、共通のポイントカードを使用している場合、ポイントカード情報を加盟企業間で利用してもよいのか、こうしたポイントカードシステムを「社内」としてみなすのか、それとも第三者提供とみなすのか、こうした点において法令やガイドラインが追いついていない現状がある。

これは今後の展望にも重なるが、グーグルが今後、グーグルが作り出したユーザーの広告に最適化されたビッグデータを売買するといった事業を行う場合、避けられないのが個人情報の第三者提供である。グーグル内でのユーザーデータの利用は利用規約によって許諾がとられているため全く問題はないが、それを第三者に提供する場合にはあらかじめ許諾を得るか、個人情報を個人が識別できない形にまで適切に処理してから提供することが求められる。こうした対策をしっかりと行うのかという視点で、グーグルが今後どのようなサービス展開を行うのかを注視していきたい。

#### 4.4 その他の問題点

では最後に、グーグルのビッグデータを利活用したサービスが浮き彫りにしたその他の問題点について考察する。

##### 4.4.1 監視への反発

グーグルの強みや特徴であるオープン性(拡張性のほうがいいかも)だが、一方で「グーグル八分」という言葉もあるように、専制的ともとられることがある。その理由のひとつは、グーグルがあまりにもシェアを拡大したため、グーグルの縄索アルゴリズムは非公開な状況の中でグーグルにとって都合の悪い情報は検索から省かれるのではないかとこの恐怖や

疑心を抱かせてしまっていることにあると指摘する意見もある(現代思想 P43 索引)。

特にアメリカなど欧米諸国では、自由や平等といった権利に対する意識が高いため「政府が個人の行動を監視する」ということに対する反発が強い。昨今でもエドワード・スノーデン氏による政府告発事件<sup>22</sup>など、政府による監視への反発に関連した出来事が発生している。一部の人々はグーグルがこうした政府の監視に協力しているのではないかと批判している。

#### 4.4.2 忘れられる権利

また、ヨーロッパでは「忘れられる権利」を法制化しようという動きがある。ウェブ上に一度公開した情報というのは、デジタルデータの複製容易性などから無限にコピーされ続け、完全に消し去ることは技術的に非常に困難といわれている。一度ウェブ上に載せてしまった個人情報、本人の望まない形で世界中にさらされ続けることに対して、こうした情報の削除を求めることができる権利を「忘れられる権利」としてEU(ヨーロッパ連合)が認め、個人データ管理者はデータ元の個人の請求があった場合に当該データの削除が義務付けられた。2014年5月には、欧州連合(EU)司法裁判所は人には「忘れられる権利」があると判断した。自分の過去の債務記録へのリンクの削除を求めていたスペイン人男性が勝訴し、グーグルはリンクを削除することとなった。欧州委員会のビビアンヌ・レディング副委員長は「忘れられる権利は人々を守り、ネットへの信頼を高めることに繋がる。」と語っている。

こうしたプライバシー問題に関して、グーグルの立場は基本的には次のようなものである。まず、メールや検索・閲覧履歴といった情報は、人間の手を介さずにデータベース化されている。よって、利用者が懸念しているようなプライバシーの侵害は起こりえない。また、グーグルは単にウェブにあった情報を整理して提供しているにすぎないという立場をとっている。検索エンジンがもたらす利便性には代償が伴う可能性についても言及しており、「今後、人々がネット上に何かを掲載する際にはグーグルが存在する限り、『この文章は永久に自分について回る』可能性があることを考慮する必要がある」<sup>23</sup>としている。

#### 4.4.3 データ収集への嫌悪感

さらに、自分の個人的なデータが収集され、それをもとにサービスが展開されることに対して嫌悪感を抱く人もいる。「グーグルナウ」というサービスがある。このサービスでは、グーグルカレンダー、Gメール、GPS位置情報、グーグルマップ情報など様々なサービスを連携させ、ユーザーがいま最も知りたいと思われる情報を「要求される前に」表示する

---

<sup>22</sup> 中央情報局(CIA)及び国家安全保障局(NSA)の局員として、アメリカ政府による情報収集活動に関わった米国人エドワード・スノーデン氏が、2013年にNSAによる個人情報収集の手口を告発した事件。

<sup>23</sup> スティーブン・レヴィ(仲達志・池村千秋訳)『グーグル ネット覇者の真実』2011年 269頁

ことが出来る。単純なものでは、天気や気温、ニュース、株価などを一日の初めに表示することが出来る。また、現在地に近い人気のあるスポットなどをあらかじめ表示させることが出来る。さらに複雑で便利なものだと、カレンダーにミーティングの時間と場所を入れておくと、自分が今いる場所からその場所までの交通渋滞状況を考慮して、「いつもより渋滞しているため、あと何分で出発しましょう。」というアラートが出る。実際にこのサービスをアメリカ在住の日本人実業家である海部美知氏が著書の中で紹介した事例では、仕事でよく行くビルを、グーグルナウが本人のオフィスと予想して「オフィスまで今日は30分かかります。」とアラートを出してきたそうである。海部氏はこのビルをグーグルナウに仕事場とは登録していなかったが、日々の行動からサービスが推測して「オフィス」とみなしたのである。さらには、登録しておいたミーティングが終わる時間には、「家まで20分」といった表示がなされる。

このように、グーグルやビッグデータの活用には、自分のあらゆるデータを収集されそれが利用されてしまうという面があり、こうした面への不快感を持つユーザーも少なくない。しかし、実際に印象論ではなく各種意識調査等を鑑みると、人々はそれほど利用データ等を収集されるサービスを使用することに嫌悪感を抱いていないことが分かる。ここではまず、位置情報を提供する場合の意識調査を挙げ考察していく。

総務省が行った調査<sup>24</sup>では、「自分の位置情報を提供することに対する考え方」という項目において、各年代のおよそ半数以上、最も情報提供意思が高かった世代では70%以上ものユーザーが情報提供をしてもよい、情報提供はやむを得ないと回答している。その一方で、位置情報をどのように利用されるのかについて、「あなたは、自分の携帯の位置情報サービス利用・位置情報の提供の設定をどこでオン・オフにするか知っていますか」という項目によると、携帯電話の位置情報サービスのオン／オフの設定方法の認知度は、全体で39.1%と認知度自体が低く、また設定方法を知らない人も45.6%と半数近く存在した。さらに、「携帯の位置情報をオンにしたり、各種サービスへの位置情報の提供を認めると、位置情報を取得した企業等が、他の情報とともに広告、商品開発等に用いることがあります。このことを知っていますか」といった項目でも同様に認知度が低く、30.3%となっている。

この調査から、ユーザーはデータを収集されることは知っているが、データの使われ方や解除方法などを知らずにサービスを利用していることが分かる。この調査は、現状では多くのユーザーが自らのデータがどのようにサービス利活用されているかについて認知度が低い、それゆえに嫌悪感を抱くなど反発することなく使い続けているという状況を浮き彫りにしたのではないだろうか。本節ではグーグルへの嫌悪感を取り上げているが、グーグルが嫌悪感を与えるのは、グーグルユーザー自らデータへの意識を強く持っているからではないかと考える。自分が使用しているサービスで個人データがどのように使われているのかを正しく知れば、ここには使われたくない、ここには使われてもよいといった意識が生まれるはずである。こうした意識のネガティブなものが、前述のグーグルナウへの

---

<sup>24</sup> 位置情報の利用に対する意識調査 総務省 平成26年

嫌悪感だと考える。

こうしたビッグデータサービスに対して初めは嫌悪感を抱くユーザーがいる一方で、私自身は、やがて利便性により「慣れていく」ユーザーが増えるのではないかと考える。Gメールの広告でも、メールを盗み見られているといった反発があったが人が作為的にメールを閲覧することはないとする説明や、なにより使いやすくスパムメールも少ない機能が好評を博し普及を遂げた。しかし、前述の調査のように、多くのサービスではユーザーは自らの利用データについての意識が低いのが現状である。提供元や業界はユーザーへのサービス認知度の向上を目指し透明性のある運営を目指すべきである。

#### 4.4.4 技術革新かプライバシーか

このように、個人データの収集と活用にはユーザーに嫌悪感を抱かせてしまう面もあるが、その一方でビッグデータの利活用を積極的に行わないことが果たして正しいことなのだろうか。ビッグデータは第三次産業革命とも呼ばれるような技術革新の可能性を秘めているとも言われている。グーグルの事業はすべてがビッグデータであり、人類史最大規模のデータベースになるかもしれない。このような技術革新という点からみると、基本的には「データを集める」部分に制約を設けるべきではない。どんなデータも目的によっては有用なデータになり得る可能性があり、多様なデータが利活用されることでそれまで存在しなかった新たな革新性や有用性が生まれることもある。私自身は技術革新という面からデータの利活用は推進するべきだと考える。しかしそのためには、サービスプロバイダはデータ利活用前にユーザーが不快感を持たないよう、データの利活用の目的や方法を認知させることが必要不可欠である。

## 5 展望

---

最終章では、今後のグーグルによるビッグデータを用いた、すでに開発が進んでいるものや今後登場すると予想されるサービスを有用性や問題点などから考察していきたい。

### 5.1 今後のサービス展開

#### 5.1.1 グーグル・カー

まず取り挙げたいのが、グーグル製の自動車(ここでは便宜上、「グーグル・カー」と呼ぶ)である。グーグルは2013年に実験的自動運転車として「lidar(ライダー)」とよばれるレーザー画像による検出及び測距技術を内蔵したオートパイロット(無人)自動車でカリフォルニアなどの公道80万キロを走行し、わずか一件軽微な擦り傷がついただけであったという報告を出した。このグーグル・カーはその後も改良が進められ、目の不自由な男性を乗せたテストでも安全性を示した。さらに2014年12月、ロイター通信はグーグルがアンドロイドOSを自動車向けにも開発すると発表したことを報じている<sup>25</sup>。すでにスマートフォンを車に接続することで一部の機能を車内で使う技術は開発されており、来年初めにも製品が発表されるとみられている。グーグルはこれをさらに進め、スマホをつながなくても車自体がスマホと同じような機能を持つようになり、地図やナビゲーション、アプリなどが使えるようになると思われる。ロイター通信は「利用者の行動範囲、使ったガソリンの量、立ち寄った場所など、広告業界がほしがる貴重なデータを収集することになる」と指摘している。さらに現在、グーグルは自動車事業を共に推進するパートナー企業を模索しており、積極的に車産業からの協力を得ることで市場の開拓と独占を狙うと考えられる。

こうしたグーグル・カーが実用段階に入った場合、どのようなことが考えられるだろうか。

まず考えられるメリットとして、交通事故の減少が見込まれるだろう。自動車がまさに「自動」で動作するようになり、グーグル・カー同士で連携が取れるようになれば、居眠りなどの人為的な操作ミスによる事故は皆無になるだろう。しかしこの実現にはすべての車両にアンドロイドなどの共通OSや制御ソフトウェアが搭載される必要があるため、各社が競争し乱立する自動車業界の現状や、一企業の独占を禁止する世界各国の経済政策を鑑みると実現は難しいかもしれない。また、日本の自動車メーカーの業績や市場拡大の追い風になる可能性もある。グーグルは自社の自動車開発においてパートナーを模索していることを述べたが、このパートナーにトヨタやホンダなどの日本企業が協力できれば、世界中のグーグル・カーに日本企業の製品が搭載される可能性もある。

一方のデメリットを考察すると、まず考えられるのはOSがウイルス等に感染し危険運転や事故を引き起こす可能性であり、安全性というメリットの影の側面でもある。しかし現

---

<sup>25</sup>朝日新聞 夕刊 2014年12月19日

状では、ロイター通信の報じた通り制御システムではなくカーナビゲーションシステムのような機能を搭載することを目指して開発されているとみられるため、こうした問題は起きにくいのではないかと考えられる。現状でも衝突回避や追い越し禁止、さらには道路の速度制限を自動で識別する機能を持つ車両も開発されているため、安全性の面よりもナビゲーションにおけるサジェスト機能などの面でグーグルは自動車産業に挑戦しようとしているのではないだろうか。

### 5.1.2 グーグル広告

これからのグーグルを考えるうえで欠かすことが出来ないのは、収益源である広告である。この収益源が今後どのような形に発展していくのか。従来のウェブ型、つまり検索ワードに関連した広告を広告主から募るという方法は、検索という手法が廃れない限り今後も主要な存在であるといえよう。そしてさらに発展してくると考えるのが、アンドロイドを搭載した端末における広告であると私は考える。日本のウェブ利用調査会社「ニールセン」による調査では、ウェブへのアクセスに際してコンピューターからのアクセスよりもモバイル端末からのアクセスが上回ったという結果が表れている。このことが示すのは、これからのウェブ利用は、モバイル端末において成長が期待されるということである。

ではモバイル端末上での広告とは一体どのようなものなのだろうか。ここからの考察はいささかSFめいてしまうきらいもあるが、お付き合い願いたい。

まず、アンドロイド端末で収集される様々なシグナルによって、ユーザーの行動や嗜好がより実生活に近づく形で把握される。例えば、ユーザーが地図上のどのルートを利用して通勤し、食事には何をどの程度どのタイミングで楽しみ(これは、クレジットカード情報アカウントと紐付けられれば把握可能。それにはグーグル社のクレジットカードが必要になるだろう...)、仕事を終えた週末にはどの店で服や化粧品を購入するのか、行きつけのテニススクールやゴルフ場の場所、もし晴れならば、雨ならば...。このようなことまでが把握可能になり、それらのデータを利活用してユーザーの携帯する端末に広告を表示させることも可能になるはずである。スマートフォンやウェアラブル端末だけでなく、車や家電、街中のデジタルサイネージ<sup>26</sup>なども連携すれば、今までにないピンポイントなマーケティングが可能になる。2002年の米映画「マイノリティ・リポート」では、近未来の日常風景として、街角に設置されたカメラが人々の網膜をスキャンし、個人に合わせた広告を展開するシーンがある。あるシーンでは、トム・クルーズが演じる主人公のジョン・アンダートンが悲壮感漂わせた顔つきで歩いていると、建物のガラス壁にアメリカンエクスプレスの広告で”Get away, John Anderton. Forget your troubles.“ (ジョン・アンダートン、旅に出ましょう。)といった表示がなされる。主人公の表情を読み取り、本人を名指しで広告を打つシーンである。21世紀初頭のSF映画のワンシーンが実現する世界は、すぐそこにや

---

<sup>26</sup> 商業施設や公共空間などで、ネットワークに接続したディスプレイで映像や情報を表示するシステムのこと。

ってきている。

その他にも広告に使う個人データそのものをグーグルが販売するようになることも考えられるかもしれない。しかし、私自身はこの可能性は非常に低いと考える。なぜなら、グーグルは顧客データを売るために事業を行っているわけではなく、世界中の情報を誰しもがアクセスできるように整理することが目的であるはずだ。整理して得たデータを売り利益を得るということは理念に反すると考える。

### 5.1.3 遺伝子情報、人工知能

また、グーグルのビッグデータの今後を考えた時に懸念される問題がある。それは人の「遺伝子情報」のデータベース化に関する問題である。グーグル創業者のサーゲイ・ブリン氏の妻が創業した「23 アンド・ミー」というバイオテクノロジー企業では、唾液サンプルを送ると個人の遺伝子をスキャンし、遺伝的な健康上のリスクなどの解析結果を知ることが出来るサービスを提供している。このサービスは直接グーグルが関わっているわけではないものの、「すべての情報を整理する」グーグルであればそこに遺伝子情報が含まれていてもおかしくはない。こうしたサービスが普及した暁には、本人でも知り得なかった遺伝子情報を第三者が収集・管理することに対する反発が生まれるだろう。サービスとプライバシーのバランスがとれるとは考えにくい。また、本人が死去した後の遺伝子データとは、その後消去されてまったく活用されないのだろうか。プライバシーを考慮してこのサービスを運用する場合、そのようにするのが望ましいが、一方で研究対象として活用することでより医学や科学の進歩に貢献するデータ群であるとも考えられる。こうしたデータ活用の面でも、これまでにない問題を孕み得るのがこの分野におけるビッグデータの利活用である。

## 5.2 データの利活用を巡る対応

展望の最後として、ビッグデータを巡る企業間の相違について触れておきたい。これまでは、ビッグデータの利活用はグーグルやその他の企業で推進されており、非常に大きな武器となることを論じ、私自身もこうしたデータの利活用は推進するべきとしたが、今後、こうした時代の流れに反して、データの利活用をあえて行わないことをひとつの価値としてユーザーに売り込む企業も現れるのではないかという意見もある。以下米アップルの事例を取り上げる。

米アップルはすでに米国で始まっている決済サービス「アップルペイ」の方針として「顧客データを集めない」ことを公言した。アップルは故スティーブ・ジョブズの在命中から、スマートフォン「iPhone」のアプリなどにしても分析用のデータの収集に関しては厳しい方針を何度か打ち出している。前述のように、Gメールは迷惑メールを振り分けたり、大容量で高機能なメールサービスを支える広告を的確に表示したりするために、メールの内

容を機械的に分析している。しかしアップルは、迷惑メールの振り分けはユーザーのパソコンのメールソフトに任せて「のぞき見」はしない方針だ。チャット型のメッセージ交換にしても、グーグルは検索性を重視して履歴をGメールに統合するが、アップルの「iメッセージ」は途中の通信を暗号化している。

位置情報などの個人情報をスマホに与えることでユーザーは利便性を享受できる。だが、プライバシーと便利さのバランスの基準は人によってそれぞれだろう。

I T系企業は、これまで一方的に個人情報を集めるばかりだったが、あえてそれをしないことをセールスポイントにするやり方もある。

(参考：日本経済新聞<sup>27)</sup>)

本論文では特にグーグルの強みをビッグデータの利活用にあるとしてきたが、このようにユーザーデータを収集しないことをひとつの価値として提示する企業が支持を得る可能性もある。私自身としては、このような戦略は事業全てで行われるのではなく、部分的な側面に限られるのではないかと考える。なぜなら、ユーザーのビッグデータ分析そのものを放棄することはサービス拡充の機会を逸することにもなり大きなリスクとなるからである。今回のアップルの事例でも「メールの内容」を暗号化しているため、今後はユーザーのプライバシーが高く反映される部分では他社でもアップルの戦略がとられることが多いと考えられる。

---

<sup>27</sup>日経産業新聞 新風 Silicon Valley 林 信行(I Tジャーナリスト/コンサルタント)  
2014年12月21日

## 結びに代えて

---

これまでグーグルという企業をビッグデータという観点から考察してきた。本論文の結論としては、グーグルとは企業理念の達成に向けて、ビッグデータという強力な武器を有し、同業種の先駆者たちに追いつき追い越さんとする力を持つ ICT 企業である。また、サービスの多角性、それにより収集できるデータの豊富さ、データの収集・蓄積・分析などのプロセスにおける技術の保有、活用範囲の広さといった側面を鑑みると、単なる検索サービス企業に留まらない、世界有数のビッグデータ企業でもある。

その一方で、ビッグデータの利活用やサービスを大規模に展開することに対しては多くの懸念や課題が存在することも明らかにした。多様なサービスを展開しユーザーとの接点が多いグーグルには、データ収集や分析などからユーザーへ嫌悪感を抱かせてしまったり、専制的と反発されることもある。特に、ビッグデータの利活用においては、第三者へのデータ提供などにおいて適切な取扱いが求められるため、特にグーグルのようなビッグデータが事業の根幹を担っている企業は細心の注意を払う必要がある。グーグルが更なる発展を遂げるためには、これらの解決を避けることは出来ないはずである。

また、グーグルの事業を支えるビッグで他の利活用による広告ビジネスモデルも、今後存続していくどうかは未知数である。日本における法改正の動きなどを鑑みると、ビジネスにおいてデータの利活用を促進し、経済成長の潤滑油とみなしている状況があり、グーグルの追い風となっている。しかし、今後ますます個人データの利活用が進めば、さらなる反発や問題が起こり得る可能性も高くなる。アップルのような同業他社が個人データの利活用を行わないという価値を提示しグーグルと差別化を図ろうとすることも考えられるため、ビジネスモデルの転換を迫られ覇権を失う可能性もある。

今後のグーグルの狙いは、ウェブ上のデータのみならず、この世に存在するあらゆる事象をもデータベース化し、世界中のビッグデータをすべて秩序立てることだと考える。それはすなわち世界中の出来事や現象、人々の行動や考えを収集・整理・管理することに他ならない。さらに、グーグルの理念が実現に近づけば近づくほど、グーグルのもたらすサービスも ICT をはじめとする様々な産業や我々の生活を変えていくに違いない。こうした未だかつてないサービスを展開していく可能性を秘めているため、今後もグーグルの動向を注視していきたい。

本論文を以て二年間のゼミナール活動が終了となる。本ゼミナールの活動において、山田正雄教授や先輩方の熱心で温かいご指導を賜り、お互いに高め合える同期や後輩にも恵まれ、非常に有意義な二年間を過ごすことが出来た。お世話になった皆様に感謝の意を表し、本論文の結びの言葉とする。

## 参考文献

---

### ≪書籍≫

- 『フリー <無料>からお金を生み出す新戦略』 クリス・アンダーソン 他 NHK 出版  
2009年
- 『クラウドの衝撃』 野村総合研究所/城田真琴 東洋経済新報社 2009年
- 『グーグル ネット覇者の真実』 スティーブン・レヴィ 他 2011年
- 『情報通信アウトロク 2013— ビッグデータが社会を変える』 情報通信総合研究所 NTT  
出版 2012年
- 『一時間でわかる 図解ビッグデータ 早わかり』 大河原克行 中経出版 2013年
- 『データサイエンティスト完全ガイド』 日経情報ストラテジー 日経 BP ムック 2013年
- 『ビッグデータの衝撃』 城田真琴 東洋経済新報社 2012年
- 『ビッグデータの正体』 ビクター・マイヤー＝ショーンベルガー 他 講談社 2013年
- 『ビッグデータの覇者たち』 海部美知 講談社現代新書 2013年
- 『IT ナビゲーター2014年版』 野村総合研究所 ICT・メディア産業コンサルティング部  
東洋経済新社 2014年
- 『コンテキストの時代ウェアラブルがもたらす次の10年』 ロバート・スコープル 他 日  
経 BP 社 2014年

### ≪新聞、雑誌≫

朝日新聞

日本経済新聞

『現代思想 一月号 第三九卷第一号』 青土社 2011年

## 参考 URL

---

グーグル(日本) <https://www.google.co.jp/intl/ja/about/>  
グーグル(米国) <http://investor.google.com/>  
総務省 <http://www.soumu.go.jp/>  
NHK ニュース web <http://www3.nhk.or.jp/news/>  
日本経済新聞 <http://www.nikkei.com/>  
ITmedia <http://www.itmedia.co.jp/>  
ウィキペディア <http://ja.wikipedia.org/>  
WIRED <http://wired.jp/>  
MSN 産経ニュース <http://sankei.jp.msn.com/>  
ITPRO by 日経コンピュータ <http://itpro.nikkeibp.co.jp/>  
ギズモード・ジャパン <http://www.gizmodo.jp/>  
Yahoo!ニュース <http://news.yahoo.co.jp/>  
CNET JAPAN <http://japan.cnet.com/>  
日経ビジネスオンライン <http://business.nikkeibp.co.jp/>  
TechCrunch Japan <http://jp.techcrunch.com/>  
Gigazine <http://gigazine.net/>

※URL は、2015 年 1 月 30 日現在のものである。