

修士論文概要書

2010 年 2 月提出

専攻名 (専門分野)	情報理工学	氏 名	櫻井 宏樹	指 導 教 員	山名 早人教授 印
研究指導名	並列・分散アーキテクチャ研究	学籍番号	CD 5108B054-6		
研 究 題 目	WWW における P3P コンパクトポリシーの利用状況に関する調査				

概要

本研究では、W3C の P3P 規格で規定されているコンパクトポリシーの利用状況について、調査を行う。コンパクトポリシーとは、サイトにおける個人に関わる情報収集に関してのポリシーを機械的に処理でき、かつ可読性を持たせた状態で公開するための設定値である。

本調査ではまず、検索エンジンを例として著名なサイトにおける設定値に関して、設定されている割合、適切に設定されているか否かを調査した。また、ランダムに収集した 846,877 件のサイトに対し同様に調査した。結果として、著名なサイトである検索エンジンでも設定されている割合は 5.69%と低く、ランダムに収集したサイトではさらに低く、0.24%のサイトしか設定していなかった。また、検索エンジンなどの著名なサイトでは適切に設定されている可能性が高いことが確認できたが、ランダムに収集したサイトでは、64%のサイトにおいて適切に設定されていない可能性が高いことが確認された。

1. はじめに

広告表示などに利用される cookie は単体では個人情報とはならないが、個人に関わる情報として、重要視され、問題となることもある。その一例として、2000 年に起きたダブルクリック社の訴訟問題[2]がある。

この問題を受けて、2002 年 4 月に W3C(World Wide Web Consortium)[3]では、P3P(Platform for Privacy Preferences)[4]を規定し、Web 上の個人に関わる情報の取り扱いに関してのポリシーを公開することを推奨するようにした。本調査で対象としているコンパクトポリシー(以下、CP)もこの規定に組み込まれているものである。また、cookie などの個人に関わる情報は個人情報保護法の保護下にないため、ポリシーの公開は義務ではない。また、内容の正確性も同様である。

ただし、Internet Explorer(以下、IE)では、デフォルトの設定の場合、CP を設定していないサイトのサードパーティ cookie を受け付けない。そのため、サードパーティ cookie を使用するために、CP を設定しているサイトは多いと考えられるが、それらの全てが CP を適切に設定しているとは限らない。

P3P に関しての調査は少なく、Cylab の調査[5]のように、ポリシーに関しての調査の一項目として触れられる程度であり、CP に関して調査したものはない。そこで本調査では、CP を主題として調査を行う。

本調査では、まず CP について、著名なサイトの例として検索エンジンの設定状況を調査した後、Web 全域で設定しているサイトがどの程度あるのか、適切に設定していないと思われるサイトがどの程度あるかを調査する。

なお、本論文では、第 2 節で実態調査の結果について報告した後、第 3 節では既存の類似調査と比較を行う。第 4 節で本調査についての考察を行い、第 6 章でまとめる。

2. 実態調査

2.1. 検索エンジンにおける CP の設定値

2.1.1. 設定されていた検索エンジンの数

Wikipedia にて著名な検索エンジンとして紹介されている 21 サイト[6]を対象に、ポリシーファイルが設定されているか、CP が設定されているかを調査した。結果として、CP・ポリシーファイル共にあったサイトが 4.19%、CP のみのサイトが 1.5%、ポリシーファイルのみのサイトが 2.9%であった。

2.1.2. 検索エンジンのドメイン間の CP の比較

ポリシーファイルもしくは CP の見つかった検索エンジンについて、jp ドメインと com ドメインで比較を行った。ただし、両ドメインで公開している検索エンジンのみを対象とする。また、参考として、世界で最も著名な検索エンジンである Google についても調査した。結果は、表 1 のようになった。

表 1 jp,com ドメイン間の比較

	jp (ポリシーファイル/CP)	com (ポリシーファイル/CP)	jpドメインと comドメインで CPが同一か
Google	無/有	無/無	
Lycos	有/有	有/有	同一
NAVER	無/有	有/有	異なる
Yahoo!	有/有	有/有	異なる

2.2. web 全域での CP の設定状況

2.2.1. 設定されているホスト数

kantei.co.jp を起点にランダムにリンクを辿り収集を行った。2009 年 3 月 27 日から 4 月 27 日の一ヶ月間収集を行い、846,877 サイトの収集を行った。そのうち、CP が設定されているサイトは 2066 サイトであった。

2.2.2. 複数のサイトで見られる設定値について

設定値を調べたところ、391 種類の設定値が見つかった。また、一部の設定値があまりに多く共通して設定されているのが見られた。本来、CP はサイトごとに自サイトで利

¹ 日本の個人情報保護法下において[1]。なお、本論文では、個人情報保護法とした場合、日本の個人情報保護法を指すものとする。また、個人情報と表記した場合は個人情報保護法において個人情報として扱われる情報を示し、個人に関わる情報と表記した場合は、必ずしも個人情報保護法において個人情報として扱われるわけではないが、個人に係る情報を示しているものとする。

² 本来であれば、適切に設定しているサイトの割合を調べるべきであるが、実際の情報の取り扱いに関してはサイト及び管理者内部ではなければわからず、外部からの調査では難しい。そのため、本調査では、適切でない可能性の高いサイトを調べるに留まっている。

用する個人情報に合わせて設定されるべき値である。これらの CP については正しく設定されたものではない可能性が高いと考えられる。これらは、「どこかからコピーしてきた。」「ツールのデフォルトの設定」の 2 点の可能性が高いと考えられる。

そこで、共通して設定される可能性が低い、policy ref を指定しているものを調査する。また、複数のサイトで設定されている設定値の情報源となった可能性があるものがあるかを調査する。

2.2.3. policy ref について

policy ref とはポリシーファイルの場所を指定するものである[4]。そのため、そのサイトに関連した URI が設定されていてしかなるべきである。URI に”http://info.yahoo.com/”と指定されている設定値からランダムに 50 件抽出し、サイト内容を調査した。結果は Yahoo! が提供しているサイトはわずか 1 サイトであった。また、Yahoo! のサービスを利用して作成したサイトが多く、設定値に関してもそのまま利用したと思われる。また、Yahoo! に関係性のないサイトに関しては、著名なサイトの例として、Yahoo! の設定値をコピーしたものだと推定される。

2.2.4. 同様の設定値となる原因

設定されていた数が上位の設定値に関して情報源を調査したところ、デフォルトの設定をそのまま使用している、設定例として公開されているものをそのまま使用しているなどの例が見られた。

また、設定しているサイトの数が多い上位 10 件の設定値について、Google にて検索をしたところ、検索結果上位 10 件のサマリーに同様の設定値が平均 5.9 件出現した。設定数が 1 件の設定値から任意に 10 個の設定値を選び、同様に検索した際は、平均で 1.4 件しか出現しなかった。

以上のことにより、同様の設定値を設定しているサイトに関しては、適切に設定していない可能性が高いと考えられる。

3. 関連調査との比較

3.1. CyLab の調査[5]

関連の調査として、Cylab の調査がある[5]。CyLab では 2007 年のプライバシーポリシーに関する調査の中で、ポリシーファイルに関しても調査を行っている。AOL, Google, Yahoo! から得られた 113880 件の検索結果(ユニークなサイト数は 80427 件)を対象に調査を行い、ポリシーファイルがあったものは 10.14% であったと報告している。

本調査における CP とは対象が違うものの、CyLab の調査では同じ P3P で規定されるポリシーファイルの普及率が低いことが見られる。

4. 調査結果のまとめ及び考察

著名なサイトの例と言える検索エンジンでさえ、あまり普及率が高くない。これは、検索エンジンではサードパーティ cookie を発行する場面が少ないことにも起因すると推測されるが、サードパーティ cookie に関係のないポリシーファイルも設定されていないことから、P3P 自体の普及があまりなされていないと考えられる。

しかしながら、設定がなされていた場合、適切に設定されている可能性は高い。Lycos の場合、参照元が全て米国のものであるため、言及は難しいが、Yahoo!

NAVER に関しては、国ごとに設定が違っており、もし適切に設定する必要がないならば、このような手間のかかることはしないと考えられるからである。

以上のことより、検索エンジンのような著名なサイトの場合、CP を設定していれば、適切に設定している可能性が高いと考えられる。

Web 全域に関しての調査では、設定しているサイトはわずか 0.24% であった。あまりに少ないことから、Web 全域での普及率も低く、設定されている例は少ないと考えられる。このような原因として、必要性のなさが可能性として最も高いと考えられる。

また、設定がしてあっても、多くのサイトが適切に設定していないと考えられる。原因として、設定値が適切であるかを確認する機構がなく、ユーザ側も注目していないことが上げられる。また、法律上の義務もないため、仮に誤った設定となっても処罰される可能性は低い。このようなことから、サードパーティ cookie を使おうとしたサイトが、「とりあえず、コピーしてしまう。」、「とりあえず、デフォルトのままで。」としてしまう場合が多く生まれると考えられる。

以上をまとめると、

- 普及率が低い
- 著名なサイトにおいては適切に設定されている可能性が高い。
- Web 全域で見ると適切に設定されていない可能性が高い。

ということがわかった。

5. おわりに

公開より約 8 年経つ P3P 及び CP だが、現状、web 全域では大半のサイトにおいて適切に利用されていないことが本調査でわかった。

今後、この規定を有効活用するためには、CP を必要とする場面を増やすことと CP に関して審査する機構を作る、もしくは、ユーザ側にしっかりと確認させる機能を織り込むことなどが考えられる。

参考文献

- [1] 個人情報保護トップページ : <http://www.caa.go.jp/seikatsu/kojin/index.html>
- [2] ダブルクリック社の情報追跡に反対運動 | WIRED VISION : <http://wiredvision.jp/archives/200002/2000020303.html>
- [3] World Wide Web Consortium (W3C) : <http://www.w3.org/>
- [4] P3P: The Platform for Privacy Preferences : <http://www.w3.org/P3P/>
- [5] Lorrie Faith Cranor, Aleecia M. McDonald, Serge Egelman and Steve Sheng : ” 2006 Privacy Policy Trends Report ” , CyLab Privacy Internet Group(2007), [http://www.chariotfire.com/pub/cpig-jan2007.pdf\(2007\)](http://www.chariotfire.com/pub/cpig-jan2007.pdf(2007))
- [6] 検索エンジン - Wikipedia : <http://ja.wikipedia.org/wiki/%E6%A4%9C%E7%B4%A2%E3%82%A8%E3%83%B3%E3%82%B8%E3%83%B3>