

都道府県 Web サイトにおいて設定されているロボット排除標準の現状

○本田正美 (Honda Masami)

Keywords : Web サイト、ロボット排除標準、都道府県、包括的対応

1 背景と目的

ロボット排除標準とは、ボットが Web サイト全体またはその一部を走査することを防ぐ規約のことを指す。Web サイトの所有者がボットに対して指示したい場合には、その Web サイトの階層のルートに robots.txt というテキストファイルを置き、そこにボットに対する指示を記載する。例えば、情報通信学会であれば、「<http://www.jsicr.jp/robots.txt>」として、テキストファイルを置く。robots.txt はボットに対して特定のファイルやディレクトリを検索しないよう指示し、サイト側の指示に従う選択をしているボットは、まず robots.txt にアクセスして指示を読み込み、その指示に従うことになる。

このロボット排除標準は自治体の Web サイトにおいても設置が広がりを見せつつある。そこで本研究では、都道府県に焦点を当てて、ロボット排除標準の設置について現状を明らかにすることを研究目的とする。

2 方法

本研究では、都道府県の Web サイトについて、ロボット排除標準設置の有無を確認する。その方法は、各都道府県のトップページの URL を取得し、その末に「/robots.txt」を追加した上でアクセスを試みることからなる。もしロボット排除標準が設置されていれば、ボットに対する何らかの指示に関する情報が表示され、なければ何らかの例外処理が返される。

探索は 2025 年 9 月 12 日と 9 月 19 日に二度実施した。これは情報の更新がなされている可能性があるゆえの措置である。

3 結果

都道府県の Web サイトを走査し、ロボット排除標準が 16 団体に設置されていることを確認した。その 16 団体のうち 14 団体は、「*」という全ボットに対して包括的に対応する設定を採用していた。個別のボットに対して指示を行う事例は少数にとどまったが、東京都が 25 個の個別のボットについての設定を行っており、次いで埼玉県と山梨県がそれぞれ 11 個の個別のボットについての設定を行っていた。特徴的な事例としては、広島県があった。同県は、全てのボットに対して不許可(Disallow)を大量に設定していた。その他の事例では、ボットに対する何らかの指示を行ったとしても、それが「不許可」を設定するということであるのは稀であった。都道府県では、ロボット排除のためにロボット排除標準が設置されているわけでは必ずしもないようである。

4、結論

2025 年 9 月時点では、日本の都道府県の Web サイトにかかわり、ロボット排除標準の設置は広がりを見せつつあるものの、広く設置が浸透するという段階にはまだ至っていないという現状が明らかとなった。

【主要参考文献】

本田正美(2022)「自治体におけるウェブアクセシビリティ方針策定と年次試験結果の公開」『情報知識学会誌』32(2)、pp.197-201.