

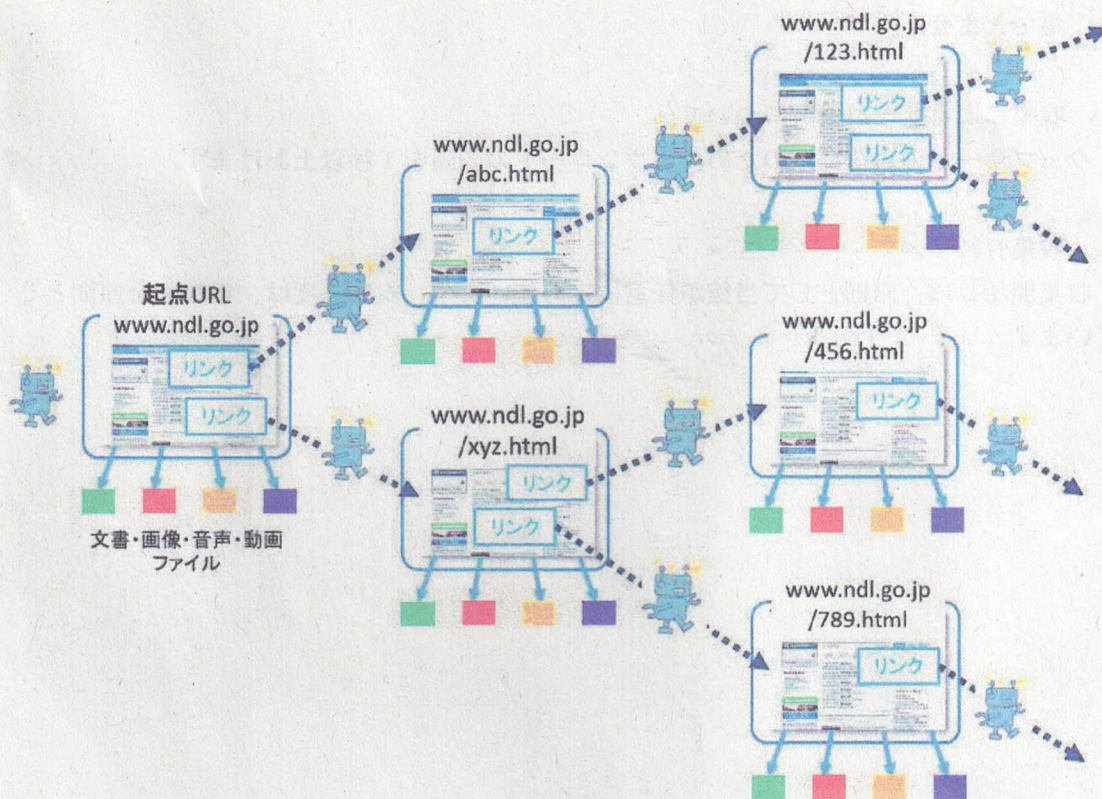
インターネット資料収集保存事業における収集の仕組みについて

1. 収集ロボットによる収集について

収集ロボット（クローラー）と呼ばれる自動プログラムを用いて、ウェブサイトを収集します。

最初にスタート地点のページにアクセスして、そのページ本文及びそこに掲載されている文書、画像、音声、動画等のファイルを収集します。次に、そこからリンクしているページに移行し、同様の動作を繰り返すことで、ウェブサイト全体のデータを収集します。

スタート地点のページは、当館で随時確認を行っておりますので、URLが変更になった場合でもその都度ご連絡いただく必要はございません。



2. ロボット排除設定について

当事業における収集ロボットのUser-agentの名前は「ndl-japan」です。当事業の収集ロボットは、データ収集を制御するロボット排除規約に準拠していますので、収集ロボットを排除する設定がされている場合は、ロボット排除設定の解除をお願いいたします。

なお、ロボット排除設定がされている場合、robots.txtファイル及びrobotsメタタグによるロボット排除設定にかかわらず、事前に連絡をした上で御許諾いただいた範囲内のデータを収集する場合がございます。

3. 収集しないデータについて

①御承諾いただいた範囲内のみ（指定したドメイン内のみ等）を収集しますので、その範囲を超えて外部サイトを収集することはありません。

②御承諾いただいた範囲内についても、部分的に収集から除外することができます。同封の回答書別紙でご指定ください。

③ID・パスワードで特定ユーザに対してのみアクセスを限定している部分や、組織内のネットワークでのみアクセス可能なデータは、収集いたしません。

④CGIプログラム等により動的に生成されるデータは収集できない場合があります。また、ダウンロードと同時に再生されるファイル（ストリーミング配信）は技術的に収集できません。

4. 収集によるサーバ負荷について

ウェブサーバの負荷軽減のため、ダウンロードの間隔を1秒以上あけます。

5. 収集日について

収集開始日は、原則として当館が任意に定めています。収集頻度は、年数回を原則としています。