

代表的な透明テキスト付 PDF ファイルの作成方法および確認方法について

【透明テキストとは】

PDF ファイルに埋め込まれたテキストデータのこと。OCR ソフトで作成されたテキストデータは、PDF ファイルの中に位置情報とともに埋め込まれるが、画面には文字として表示されることはない。そのため「透明テキスト」と呼ばれる。テキストデータを埋め込むことで、文字検索が可能になる。透明テキストを埋め込むか埋め込まないかでコンテンツのヒット率が有意に変わるため、登録コンテンツには透明テキストを埋め込むことを強くお勧めする。ちなみに、Word ファイルなどから Acrobat 等の PDF 化ソフトを使用して作成する場合は、テキストデータが自動で埋め込まれる。

※Digital Repository Federation (デジタルリポジトリ連合) 用語集より

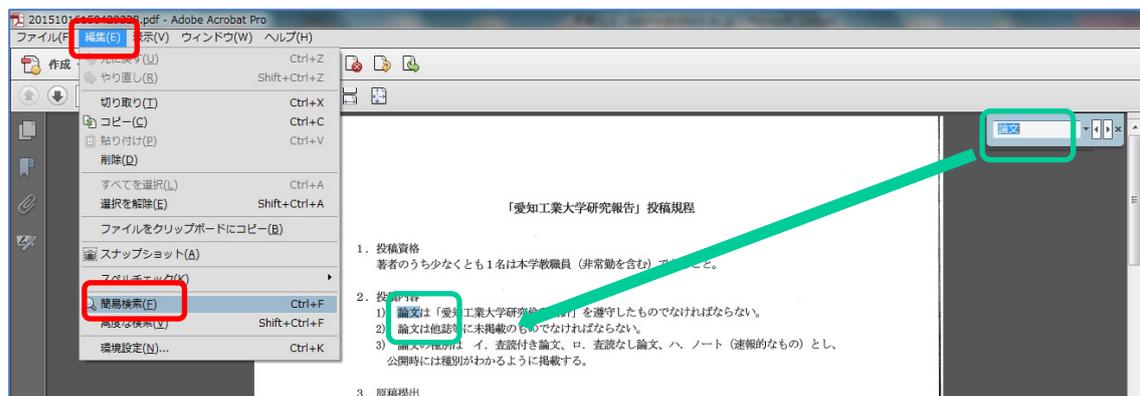
【Word で直接作成する】

※ Word で作成した論文を、「ファイル」→「印刷」→「プリンター：Microsoft Print to PDF」を選択して印刷（保存）する、または「ファイル」→「名前を付けて保存」→「ファイルの種類：PDF」を指定して保存することにより、透明テキスト付 PDF ファイルを作成することが可能です。

しかし、識字率が悪く論文中の図・表が正確に PDF 化されない場合がありますので、PDF 作成ソフト「Adobe Acrobat」(以下 Acrobat)をパソコンにインストールすることをお勧めします。

※Acrobat 導入において、ご不明なことなどがありましたら、大学計算センター（内線 1300）までお問合せ下さい。

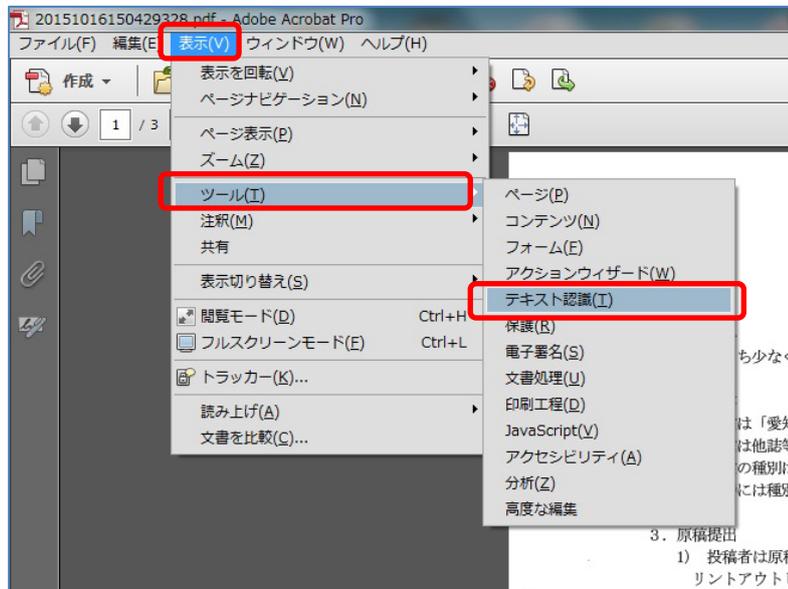
- ① Acrobat をインストールすると「ファイル」→「印刷」→「プリンター：Adobe PDF」を選択できるようになり、印刷（保存）することで透明テキスト付き PDF ファイルが作成されます。
- ② 作成した PDF ファイルを開き「編集」→「簡易検索」で論文中の単語（ひらがな、カタカナ、漢字、英字、数字などのキーワード）を入力してが検索できるか確認してください。
検索できれば透明テキスト付き PDF ファイルになっています。



【スキャナーで原稿を読み込んで作成する】

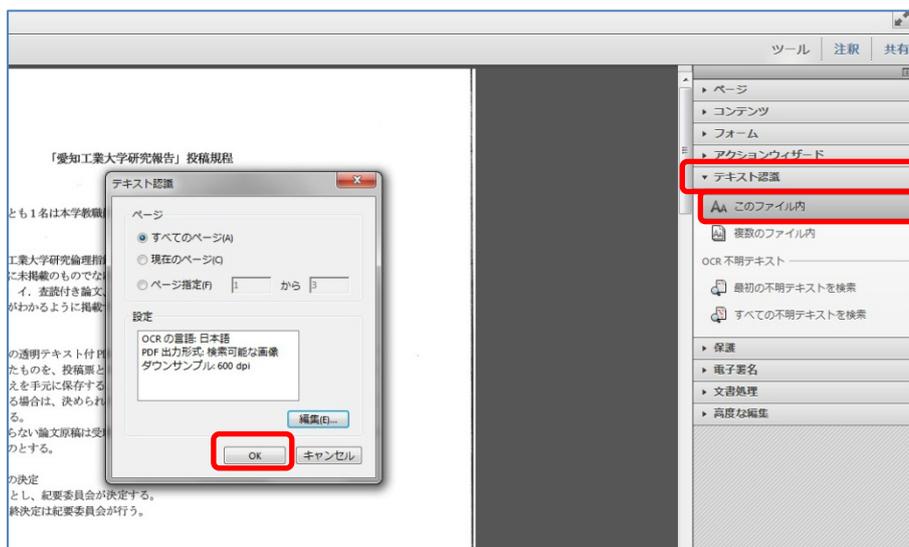
※ スキャナーを利用する場合、(解像度は 600dpi が望ましい) の A4 サイズで PDF ファイルとして読み込んでください。なお、読み込んだ段階ではまだ透明テキストが付いていませんので、OCR ソフトを使って OCR 処理を行います。ここでは一般的な OCR ソフトである Acrobat を使った処理を説明します。

- ① Acrobat を起動し、スキャナーで取り込んだ PDF ファイルを開き、「表示」→「ツール」→「テキスト認識」を選択



- ② 「テキスト認識」内の「このファイル内」を選択し、「テキスト認識」のポップアップ内の「OK」を選択

→ 透明テキスト付 PDF を作成し始めます。(「設定」はデフォルトのままで OK)



- ③ 【Word で直接作成する】の②と同じ手順で透明テキスト付き PDF ファイルになっているか確認してください。