

《実習① 雑誌(冊子)電子化実習・作業手順》

＜①書誌情報の作成・電子化整理番号の付与＞

・この実習では扱いません。

＜②背の裁断(または、コピー)＞

- ・資料の背の部分を裁断機で裁断します。
- ・事情により裁断できない場合には、裁断せずにコピーして、そのコピーから電子化する場合があります。

＜③スキャナ連続読み込み＞

操作

デスクトップ上にある次のアイコンをダブルクリックし、プログラムを起動します。



つづいて、一次情報入力システムのメインメニューで、「入力作業」をクリックします。



操作

電子化整理番号を入力し、＜決定＞ボタンをクリックするか、エンターキーを押下します。書誌情報を確認してから作業を開始します(電子化整理番号はバーコードの7桁の番号です)。



操作

メニュー<入力作業>をクリックすると、次の画面が表示されます。



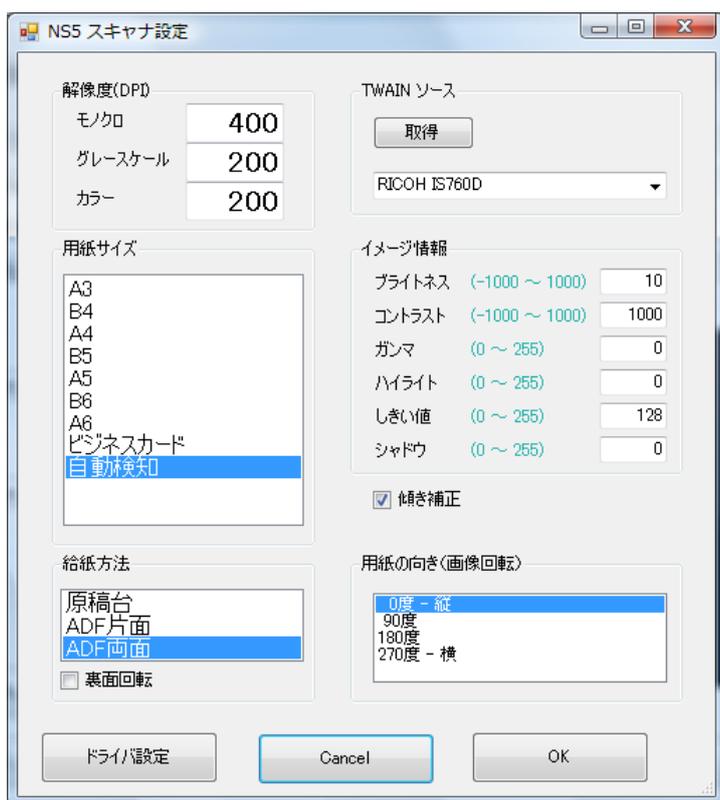
- 全頁をモノクロで、オートドキュメントフィーダ(ADF)により両面を同時に読み込んで、本文のイメージデータ(画像データ)を作成します。ここで作成されるのは、TIFFフォーマットのイメージデータです。また、スキャナ読み込みと同時に、OCR変換を行うことにより、テキストデータが作成されるとともに、

TIFFフォーマットのイメージデータとテキストデータから、頁単位の透明テキスト付きPDFが作成されます。

※ほとんどがカラーページである雑誌など、全頁を「カラー」あるいは「グレースケール」で読み込んだほうが効率的な場合もあります(この場合、イメージデータは PNG フォーマットとなります)。

操作

＜スキャナ設定＞ボタンをクリックすると次のダイアログが表示され、各種の調整ができます。主に用紙サイズを調整する場合などに使用します。
設定が完了したら、＜OK＞ボタンをクリックします。



[機能解説]

スキャナ設定 ...読み取り時のスキャナ設定を行います。

解像度(DPI)

用紙サイズ (A3、A4、A5、A6、B4、B5、B6、ビジネスカード、自動検知)

給紙方式 (ADF 片面、ADF 両面、原稿台)

イメージ情報 (ブライトネス等)

用紙の向き (画像回転)

操作

スキャナにおもて面を上にして裁断した冊子を設置し、＜スキャナ連続読み込み／OCR 連動＞ボタンをクリックするとスキャナ読み込みが始まります。

読み込みが終わったら、<終了>ボタンをクリックします。

<④OCR 変換>

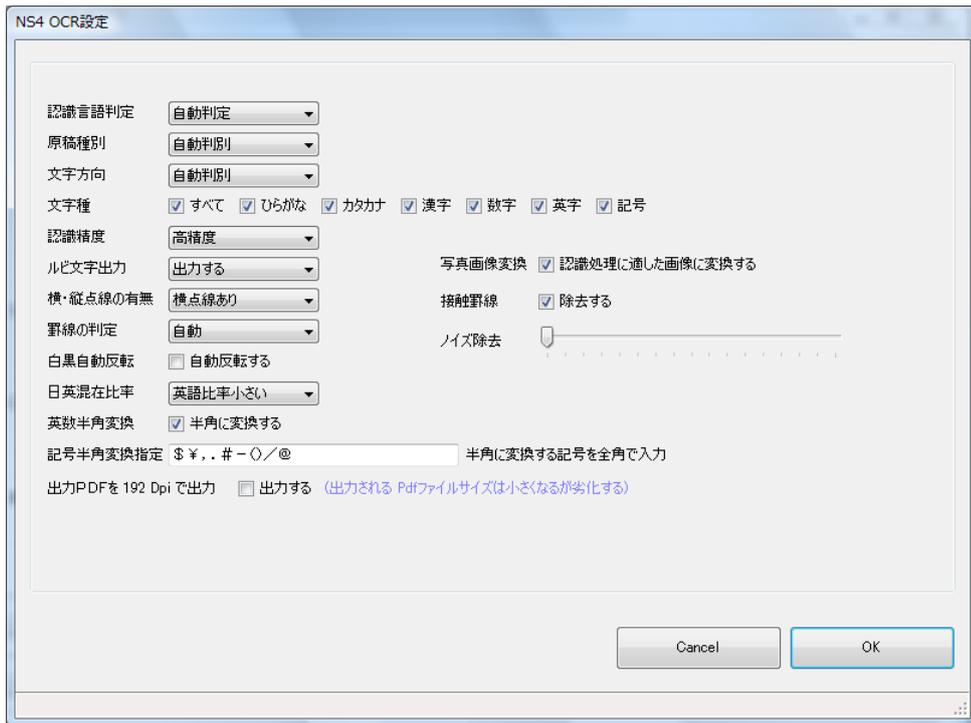
※ 入力作業で<スキャナ連続読み込み/OCR 連動>を実行した場合は、この作業をする必要はありません。

・本文のテキストデータを作成するとともに、イメージデータとテキストデータから、頁単位の透明テキスト付き PDF を作成します。

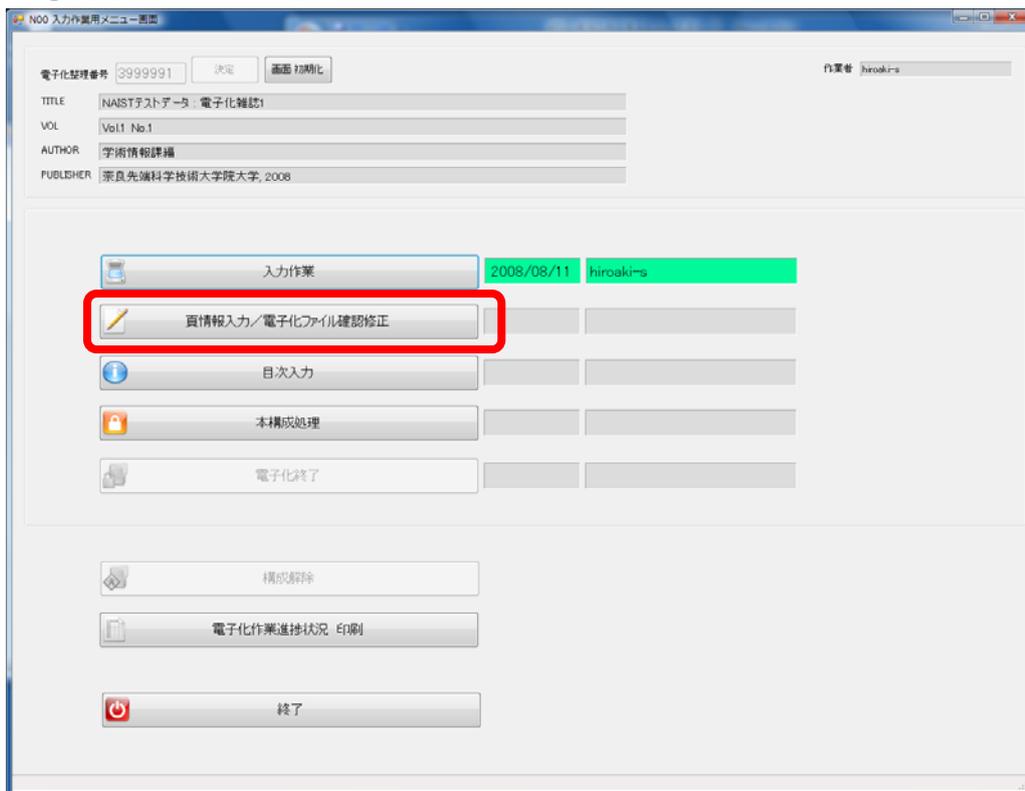
[機能解説]

OCR 設定 ……OCR の設定をします

特に洋雑誌の場合には、設定をチェックすること。認識言語判定を「英語」に、また、日英混在比率を「英語比率高い」に設定できる。



<⑤頁割付、データ確認>



操作

<頁情報入力/電子化ファイル確認修正>メニューをクリックすると、次のような画面が表示されます。



※以下、目次入力まで、各画面の右下に「頁保存」「終了」ボタンが表示されます。変更した内容を「頁保存」ボタンで保存することができます。作業を終了しメインメニューに戻る場合は「終了」ボタンをクリックしてください。

- ・スキャナ読み込みで作成した頁単位のイメージデータに、実頁としての数字を割り付けます。
- ・これは、最終的に、頁単位の透明テキスト付きPDF を論文単位または章単位のPDFにまとめるために、さらには、目次から各論文または各章へリンクするために、不可欠な情報です。
- ・頁を割り付ける際には、大きく4種類(表紙、目次、本文、頁のついていない部分)に分けます。
- ・表紙と目次には実頁をつけません。表紙には”H1”、目次には”M1”と、それぞれアルファベットをつけた数字を順次割り付けます。
- ・表紙、目次以外の頁については、実頁で割り付けます。
- ・最後に、数字のついていない頁に先頭から順に”E”をつけた数字を割り付けます。

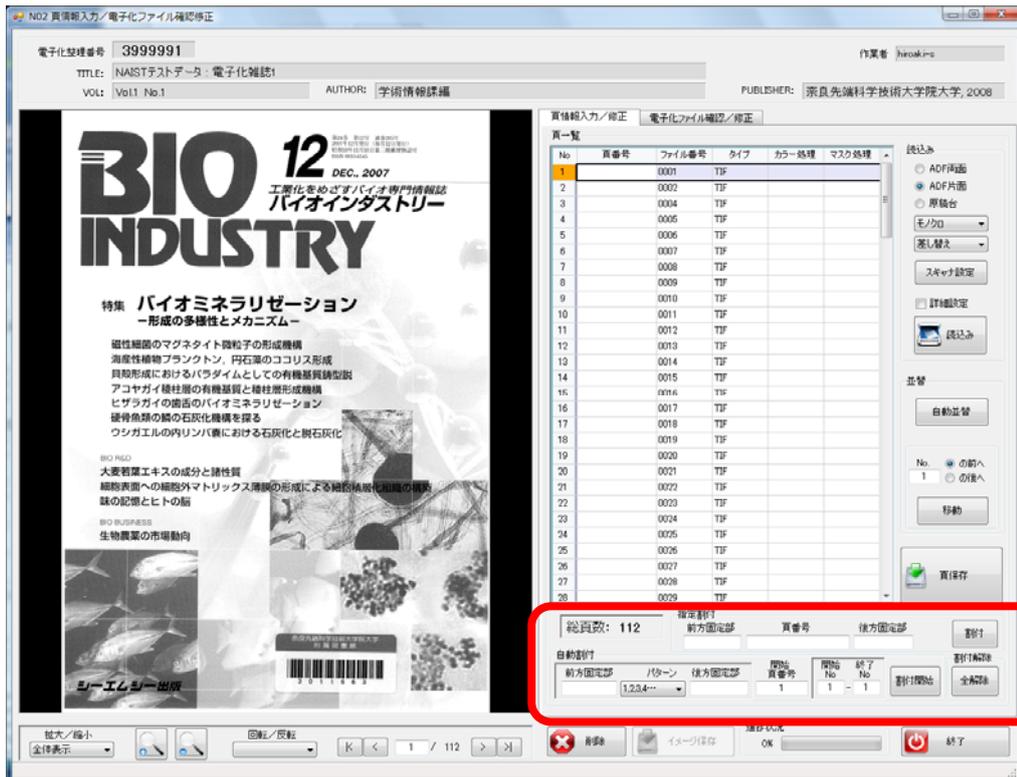
-----例-----

H1 (割り付けた頁)	1 (イメージの枚数)
E1	2
M1	3
M2	4
E2	5
3	6
4	7
5	8
E3	64

- ・ 例外として、ローマ数字(i,ii,iii…)、アルファベット付き数字(S1、L1 等)などもありますが、目次入力とそろえて割り付けすればリンクすることが可能です。

操作

画面右の頁一覧から、番号を振りたいページを選択します。必要に応じて＜前方固定部＞にアルファベットを入力して、開始頁番号を入力し＜割付開始＞ボタンをクリックすると、連続した頁番号が割付されます。その後は次の機能解説を参照しながら、必要に応じて頁番号を振ってください。頁番号を入力したら、＜頁保存＞ボタンをクリックし、ページ情報を保存します。



[機能解説]

自動並替・・・読み込む段階で作業の都合上、頁が前後している場合などにイメージの順序を並び替えることができます(現在、この機能には不具合があります)。

開始頁番号—**開始 No.**—**終了 No.**→**割付開始** …複数頁の一括割付

前方固定部—**パターン**—**後方固定部** …アルファベット等を割り付ける時に便利

前方固定部—**頁番号**—**後方固定部**→**割付** …一頁ずつの割付

操作

＜電子化ファイル確認/修正＞タブをクリックすると、次のような画面が表示されます。



- ・ここで、頁イメージが正常に読み込まれているか、各イメージデータに対応するテキストデータが、ある程度正確に作成されているかについても確認し、問題がある頁については、再度スキャナ読み込みや **OCR実行**を行います。
- ・ また、広告頁等の不要な頁は削除します。

操作

ページ削除や再スキャンをする場合は、まず頁一覧から該当ページを選択します。



操作 ページを削除する場合は、該当ページを表示して<削除>ボタンをクリックします。

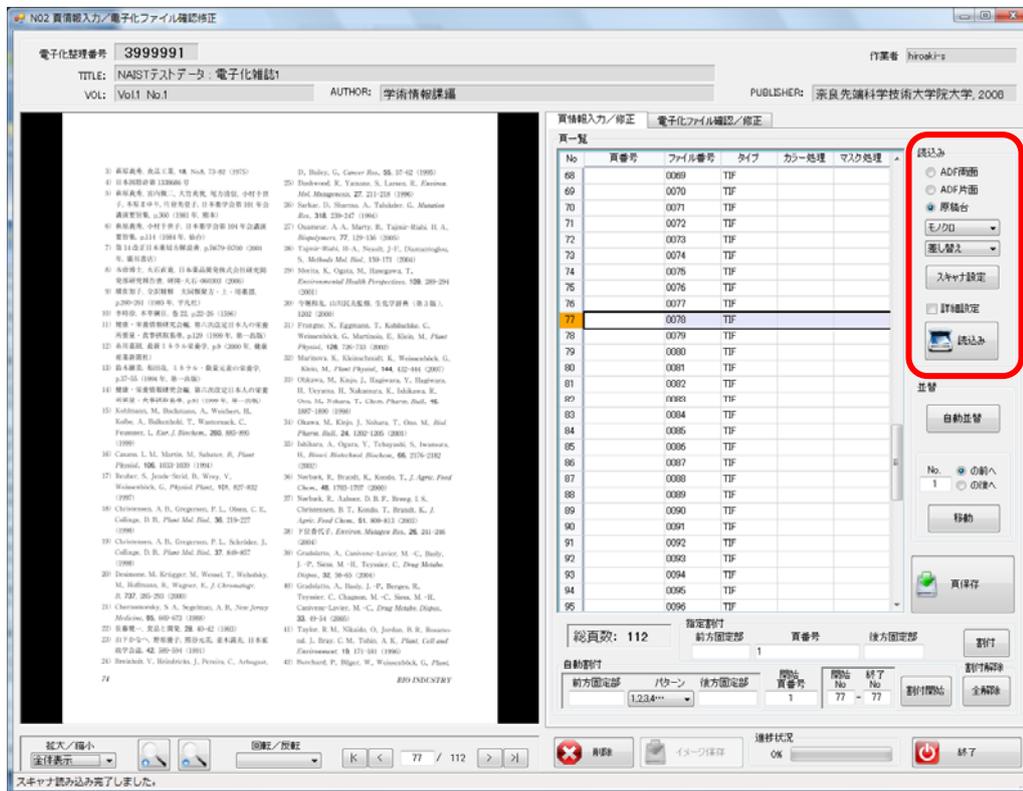


- 本文(ページ数がある場合)の白紙は削除しない。

- ・ E が付与されるページ(ページ数がない場合)が白紙の場合は削除する。

操作

ページを読み直す場合は、該当ページに移動した後、コンボボックスから<差し替え>を選択し、<スキャナ設定>(用紙サイズ等)を必ず確認して、<読み込み>ボタンをクリックします。読み込みが終わると自動的に OCR 変換も行われます。

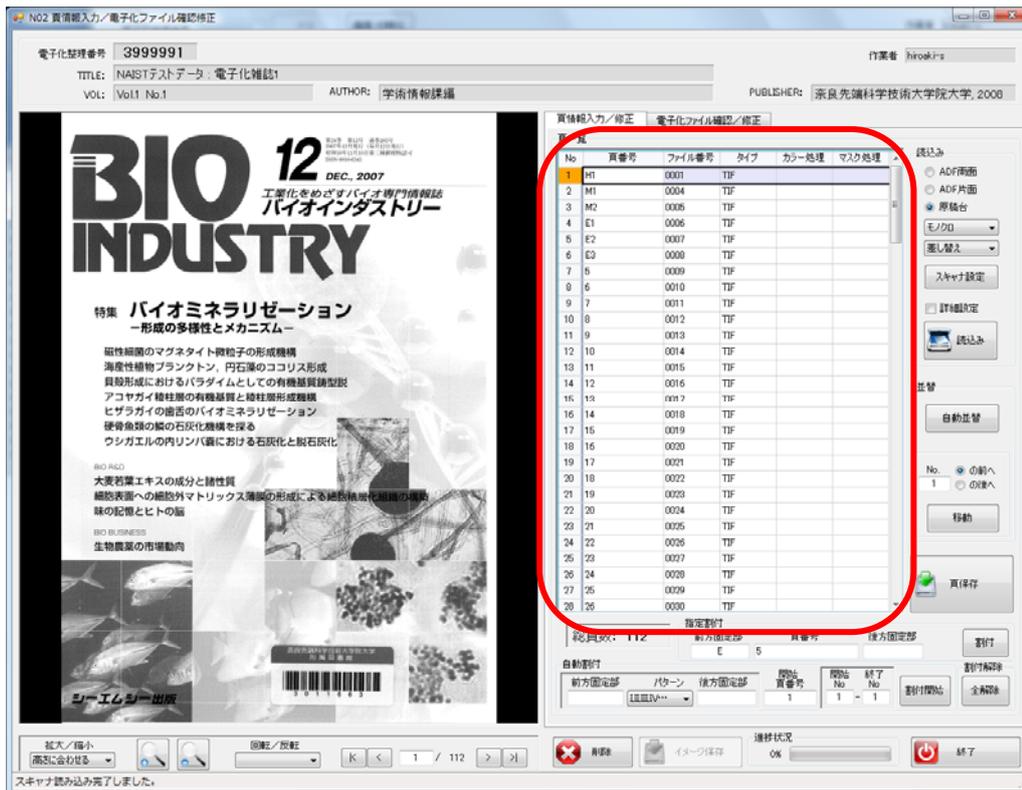


<⑥カラー頁等の再読み込み、マスク処理(OCR 変換)>

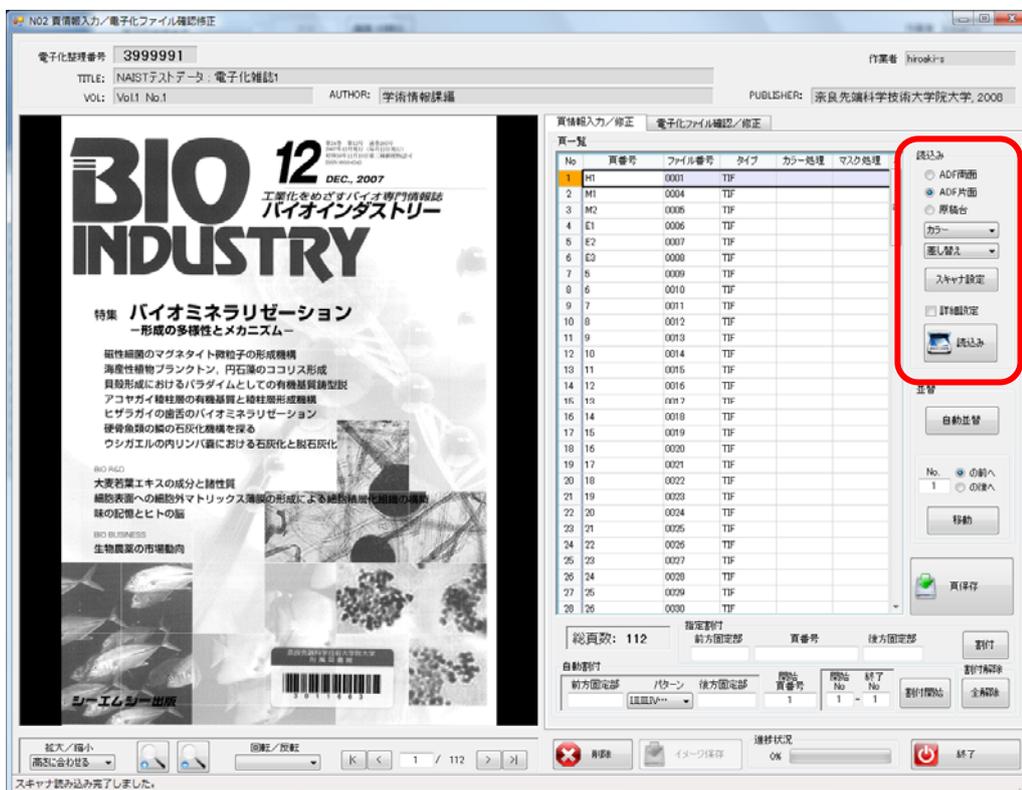
- ・白黒写真等の濃淡画像のある頁は、グレースケールで再読み込みを行います。
- ・また、カラー頁については、カラーで再読み込みします。全資料の表紙と、色付きの頁は、原則としてすべてカラー処理します。
- ・これらの再読み込みによって、PNG フォーマットのイメージデータが作成され、モノクロのイメージデータとは別に保存されます。
- ・以下では、カラー再読み込みを説明します。グレー再読み込みも操作方法は同じです。

操作

頁一覧より、カラー再読み込みしたいページをクリックして選択します。

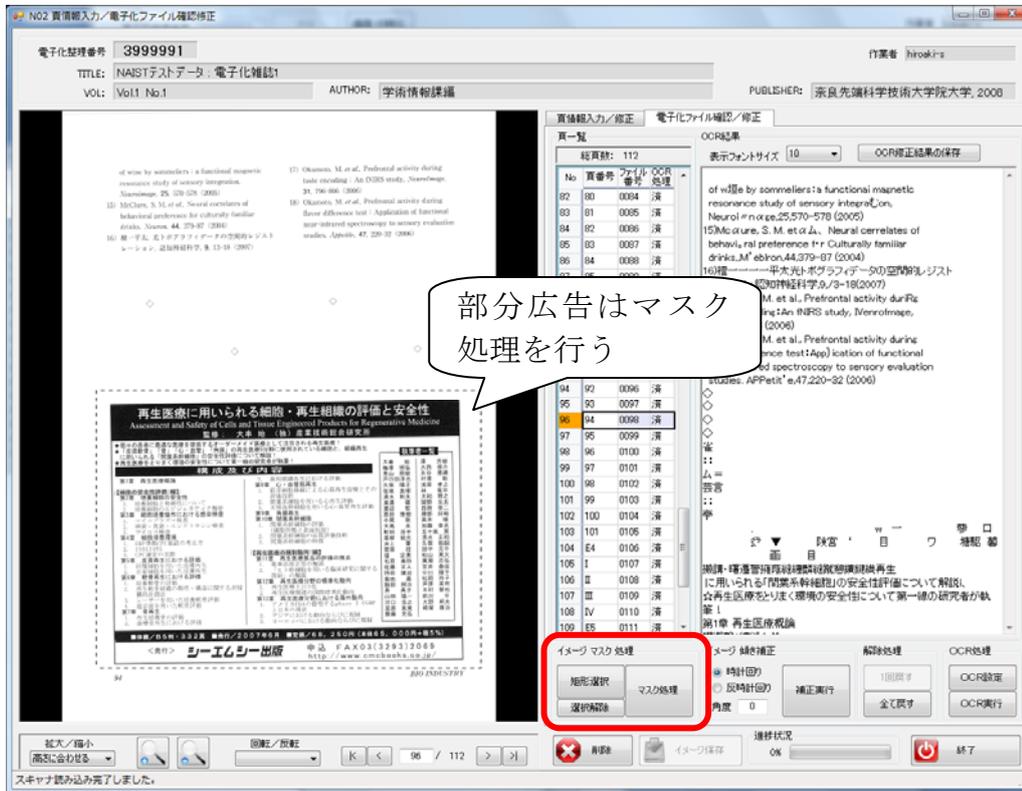


操作 モードを「カラー」にし、差し替え<スキャナ設定>ボタンで読み込みサイズなどを必ず確認し、読み込むページを上にしてスキャナに設置した後、<読み込み>ボタンをクリックします。

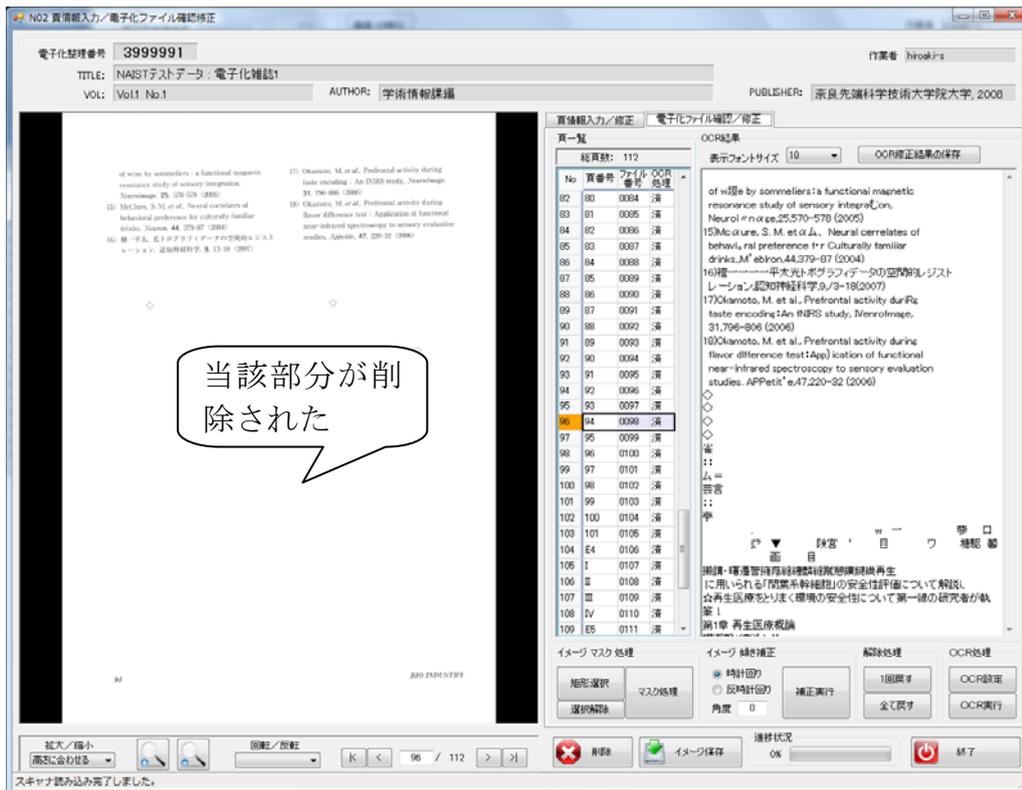


[参考]

- ・頁の一部に**第三者著作物**(利用許諾を得ている著作権者以外の著作者の著作物)が掲載されている場合には、イメージマスク処理を行って当該部分を消す作業を行います。
- ・「電子化ファイル確認/修正」タブをクリックしてから、該当頁を選択します。イメージマスク処理の**矩形選択**をクリックし、マウスで消去する部分を範囲選択します。



- ・**マスク処理**をクリックすると、該当部分が消去されるので、**イメージ保存**をクリックして、処理を完了させます。
- ※上図の例は、出版元の広告ですので、実際にはマスク処理は行いません。



・カラーやグレースケールでの頁の再読み込みやマスク処理の際にも、同時に OCR 変換を行って再度頁イメージからテキストデータを作成し、モノクロでの読み込み時に作成したものと置き換えるとともに、PNG フォーマットのイメージデータと再作成したテキストデータから、頁単位の透明テキスト付き PDF を再作成してモノクロ時のものと置き換えています。OCR 実行をクリックします。

<⑦目次入力>

- ・目次については、電子化資料をより参照しやすくするために、本文のイメージデータとは別に **テキストデータ(文字データ)** の形でも作成しています。
- ・これにより、テキストデータとして作成した目次から、本文中の目指す論文または章のイメージデータに直接リンクさせることができます。
- ・例に従って、**タイトル、著者名、頁範囲等の目次情報**を入力していきます。

[注意事項]

- ・著者名がない場合は省いても構いません。
- ・カタカナは全角に、アルファベットは半角に直します。
- ・頁範囲は半角数字、ハイフン、半角数字で範囲を示します。(例)頁範囲:1-5
- ・1頁だけの時は数字だけにします。(例)頁範囲:7

-----例-----

1)

標題: Construction of an Autonomously Replicating Plasmid in n-Ankane-Assimilating Yeast, *Candida tropicalis*.

著者名: Akihiro Hara, Mitsuyoshi Ueda, Naoki Kanayama, Atsuo Tanaka

頁範囲: 717-720

2)

標題: Nucleotide Sequences and Characterization of Genes Encoding Naphthalene Upper Pathway of *Pseudomonas aeruginosa* PaK1 and *Pseudomonas putida* OUS82.

著者名: Noboru Takizawa, Toshiya Iida, Yue-Wu Wang, Masao Fukuda, Hohzoh Kiyohara

頁範囲: 721-731

[機能解説]

目次テキスト読込 ...「メモ帳」等のテキストエディタを使って別に作成したタグ付きのテキストファイルを取り込むことができます。(P18 参照)

頁ノンブルチェック ...目次で入力した頁範囲と、頁割付で割り付けられた数字が対応しているかをチェックすることができます。

一覧 ...入力した目次情報が一覧表示されます。

OCR ...OCR 変換で作成された OCR テキストを呼び出してコピー & ペーストできます。

イメージ ...作成したイメージデータを見ながら、目次入力ができます。

- ・ 目次入力は、スキャナ読み込みの前に行うこともできます。



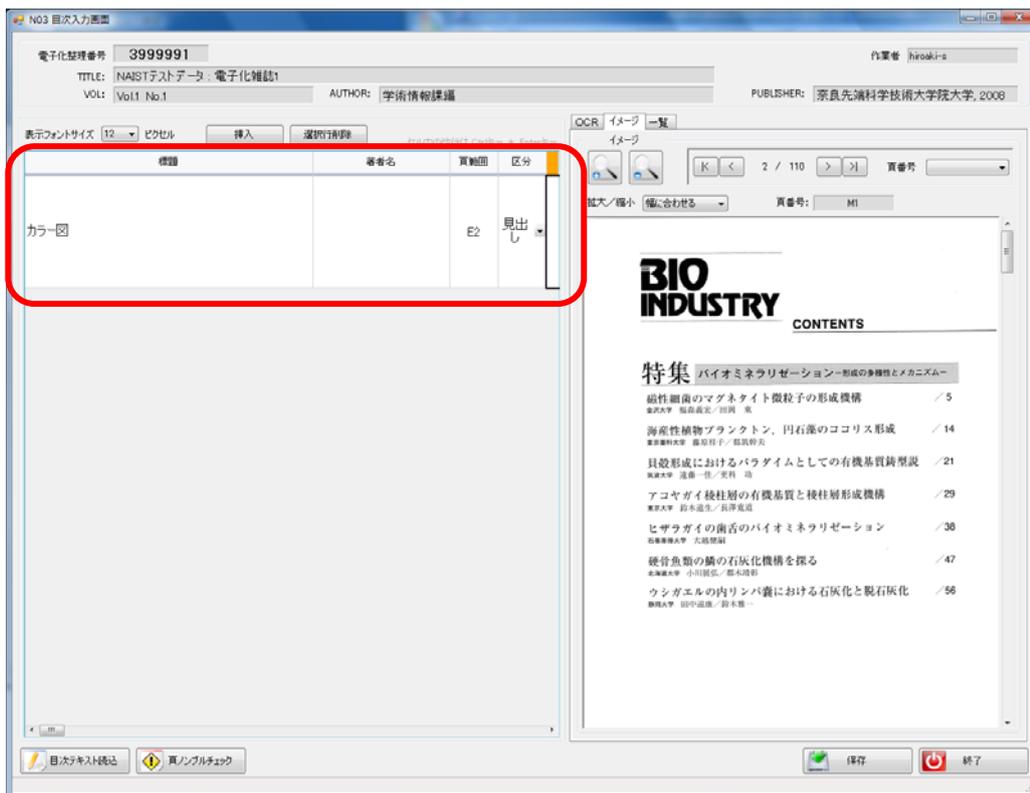
操作

＜目次入力＞メニューをクリックすると、下の画面が表示されます。



操作

前述の例に従い、目次情報を入力します。



操作

(読込んだ目次ページからコピー&ペーストする場合)

画面右の<OCR>タブをクリックし、<ページ移動>ボタンを使用して目次ページに移動すると、読込んだ目次が表示されます。適宜、OCR 結果表示領域内の文字列を範囲選択し、画面左の目次情報欄にコピーしてください。



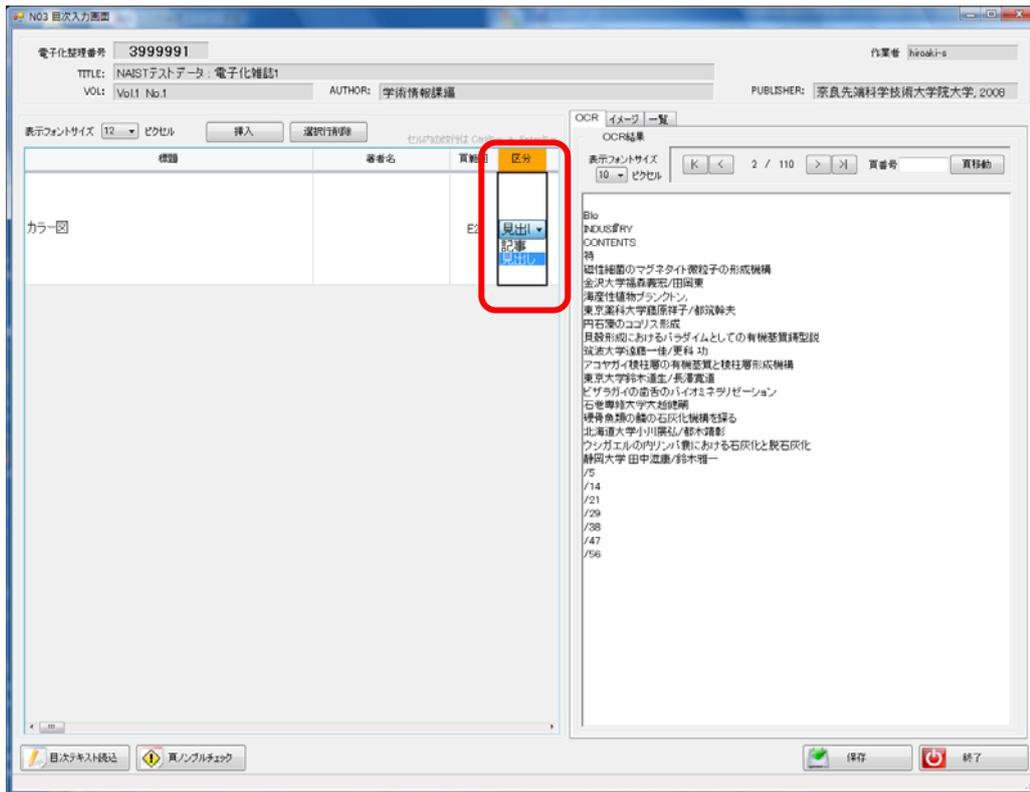
操作

一つの目次情報の入力が終わったら、頁範囲を選択した状態でエンターキーを押すと、次の目次情報入力に移ります。

題名	著者名	頁範囲	区分
カラー図		E2	記事
			記事

操作

論文以外のページについては、<区分>を「見出し」にします。(Web上でこのページへのリンクが貼られなくなります。)



[参考] 目次テキスト取込

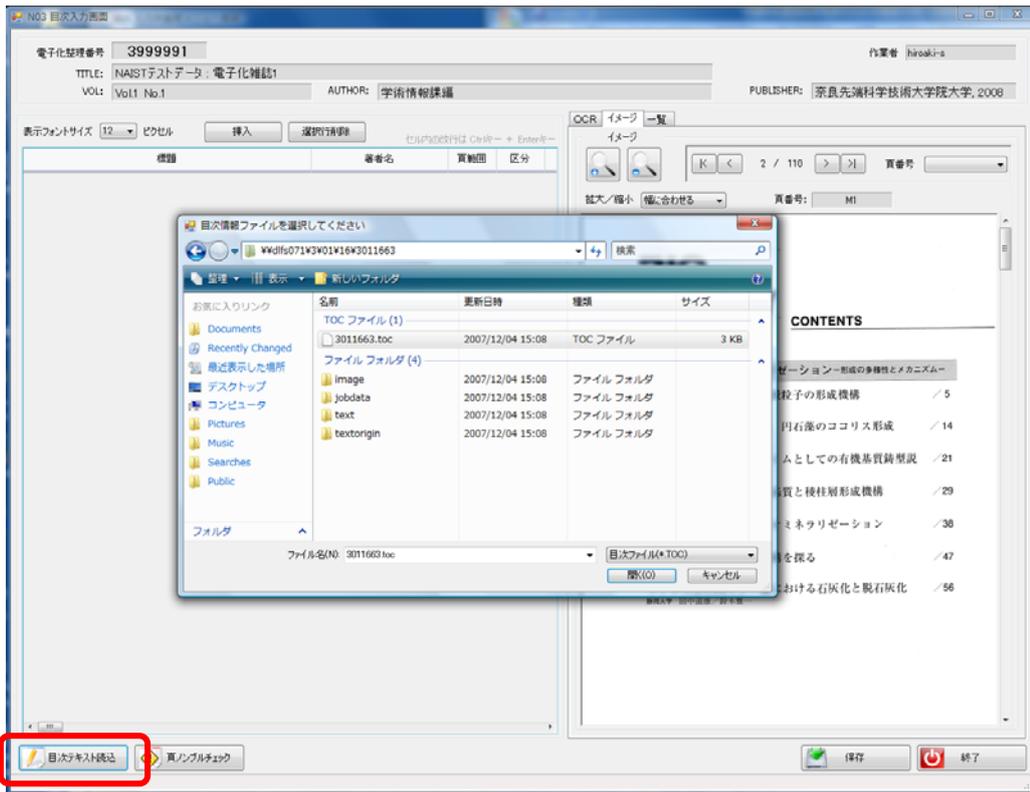
- ・テキストエディタを使用して次のような形式で目次情報を入力し、テキストファイルを作成します。

```

01 | 1 | T:特集 バイオミネラリゼーション ー形成の多様性とメカニズムー↓
02 | 2 | 磁性細菌のマグネタイト微粒子の形成機構↓
03 | 3 | R:福森義宏, 田岡 東↓
04 | 4 | P:5-13↓
05 | 5 | ↓
06 | 6 | T:↓
07 | 7 | R:↓
08 | 8 | P:↓
09 | 9 | ↓
10 |10 | T:↓
11 |11 | R:↓
12 |12 | P:↓
13 |13 | ↓
14 |14 | T:↓
15 |15 | R:↓
  
```

[T:]にはタイトル
[R:]には著者名
[P:]にはページ番号
を入力する

- ・画面左下に<目次テキスト読込>ボタンをクリックし、作成したテキストファイルを選択すると、テキストファイルから頁情報が取り込まれます。



- ・電子化作業の流れとしては、イメージデータ、テキストデータを作成して頁割付、データ確認等を行い、目次入力が終わったら、最後に本構成処理を行います。(実習では、行いません)
- ・この処理により、目次と頁割付で作成した頁情報に基づいて、頁単位の透明テキスト付き PDF が、雑誌であれば論文単位に、図書であれば章の単位に自動的に合成されて公開用の PDF が作成されます。その後、夜間処理によって、図書館システムで作成した目次情報とリンクされ、OPAC から検索、利用できるようになります。

<⑧WEB 上での確認>

- ・夜間処理によって目次情報とリンクされた電子化データ(公開用 PDF)について、利用者が実際に利用するにあたって不備がないかどうかを、WEB 上で確認します。
- ・電子図書館ホームページの「蔵書検索」(OPAC)で、該当する書誌を検索して確認します。
- ・雑誌については、「電子化雑誌一覧」の頁から確認することもできます。
- ・ビデオについては、映像や音声も確認します。
- ・問題が見つかったものについては、構成解除処理を行い、データを修正した上で、再度本構成処理を行います。
- ・確認して問題がなければ、最後に電子化終了処理を行って、電子化作業は完了することになります。

ただ、実習では、みなさんが作成したデータの確認はできませんので、既に作成されている電子化資料を OPAC や「電子化雑誌一覧」から検索していただき、電子化したものが、どのように表示されるかを見ていただきます。

操作

WEB ブラウザを立ち上げ、電子図書館ホームページ (<http://library.naist.jp/>) にアクセスします。

The screenshot shows the NAIST Digital Library homepage. The navigation menu on the left is expanded, with the following items highlighted in red:

- 情報検索 >> NAIST蔵書検索
- NAIST電子化資料 >> 電子化雑誌

The search area in the center features a search box, a search button, and a clear button. Below the search box are radio buttons for "本学所蔵", "E-Journal", and "Cross Search".

The right side of the page displays news and event headlines, including "NAIST 奈良先端科学技術大学院大学 ニュース&トピックスヘッドライン" and "NAIST 奈良先端科学技術大学院大学 イベント情報ヘッドライン".