

Neuron ES（全文検索システム）
利用者マニュアル
UI 用 v2.6

目次

第 1 章. はじめに.....	1
1.1. 対象.....	1
1.2. 制限事項.....	1
1.2.1. DocuWorks ファイルの留意事項について.....	2
第 2 章. 基本的な使い方.....	4
2.1. 検索画面へのアクセス.....	4
2.2. 検索画面について.....	4
2.2.1. 検索入力部.....	5
2.2.2. ファセット部.....	9
2.2.3. 検索結果表示部.....	11
2.2.4. プレビュー表示部.....	13

第1章 はじめに

第1章. はじめに

検索エンジンはファイルサーバーや Web コンテンツを検索できるシステムです。
利用するにあたり、その対象や注意事項、制限事項を確認ください。

1.1. 対象

対象システム

- ・ ファイルサーバー
- ・ SharePoint / SharePoint Online
- ・ Web サイト
- ・ Notes
- ・ Box
- ・ Dropbox
- ・ Google Drive、等

※上記のうち実際に対象となるシステムはお客さま環境によります。

対象ファイル

(全文検索対象)

- ・ Microsoft Office ファイル
- ・ リッチテキストファイル
- ・ テキストファイル
- ・ PDF ファイル
- ・ HTML ファイル
- ・ Visio ファイル
- ・ Open Document ファイル
- ・ DocuWorks ファイル
- ・ Google ドキュメントファイル
- ・ ZIP ファイル(内部に含まれるファイルが検索対象)

(ファイル名のみ検索対象)

- ・ 上記以外のファイルでかつ下記拡張子以外のファイル
 - ico, ini, lnk, exe, msi, dll, iso, pst, css, js, jsp, asp, aspx, java, c, pl, py, rb, cpp, cs, ps1, jar, class, bat, ini, vb, sh, php, h, vbs, cmd, dtd, scc, prj, wsdl, acx, evt, cab, db, old
- ・ 上記に該当するがファイルサイズが 20MB を超えるファイル

(サムネイル・プレビュー生成対象)

※検索エンジンにはサムネイル・プレビューを表示する機能があります。

- ・ Microsoft Office ファイル
- ・ テキストファイル
- ・ PDF ファイル
- ・ DocuWorks ファイル
- ・ 画像ファイル

1.2. 制限事項

(検索の制限)

- ・ インデックスの更新はシステム管理者が設定した間隔で更新されます。通常これは夜間となります。そのため、ファイルサーバーに保存したものがその日のうちに検索対象となることはありません。
- ・ 機種依存文字は文字化けになる場合があります。また、テキストファイルで 150 文字未満のものは文字化けする可能性があります。

第1章 はじめに

- Web コンテンツは、HTML ファイルに記載の文字コードと保存時の文字コードが異なる場合、文字化けします。
- 全文検索対象であってもパスワードが設定されている、あるいは暗号化されているファイルはファイル名のみの検索になります。
- Office XP 等古いバージョンの Office で作成されたファイルは文字化けしてしまうため、全文検索対象であってもファイル名のみの検索になる場合があります。

(サムネイル・プレビューの制限)

- PDF にコピー制限、テキスト抽出制限設定がある場合、サムネイル・プレビューが生成されません。
- 対象となる文書のグラフやレイアウトを忠実に再現できない場合があります。
- 空白ページは生成がスキップされるため、ページ数が異なる場合があります。
- ファイル名に半角&が含まれている場合、サムネイル・プレビューが生成されません。
- パスワードが設定されている、あるいは暗号化されているファイルはサムネイル・プレビューが生成されません。

1.2.1. DocuWorks ファイルの留意事項について

(DocuWorks ファイル)

DocuWorks は、MS Office や PDF と異なり、特殊な文書構成となっています。具体的には下記のような構成です。

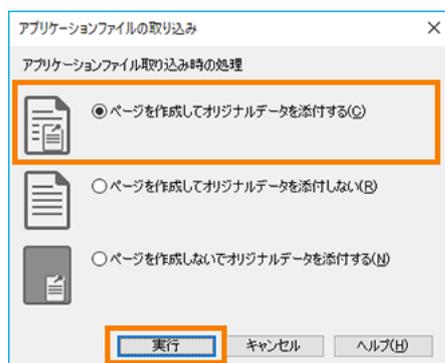


DocuWorks ファイルは、そのファイルの中に

- DocuWorks Viewer で表示されるページデータ
- 作成時に取り込んだ Office や PDF 等のオリジナル文書を含んでいます。

DocuWorks Viewer で実際に表示されるのは、ページとして作成されたデータになります。何がページとして作成され閲覧できるかは、オリジナル文書を取り込むときに指定できます。

※ https://www.fujifilm.com/fb/support/software/docuworks_91/faq/sub003-1.html



上記取り込み時の選択肢にあるように、ページを作成せずにオリジナル文書のみを DocuWorks に含むことが

第1章 はじめに

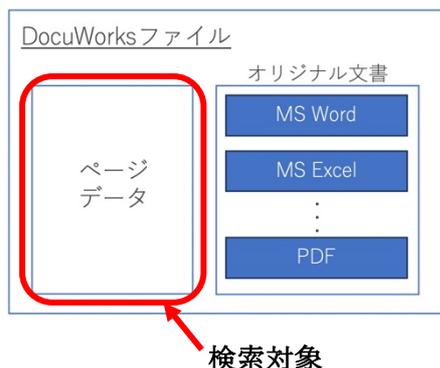
できます。この場合、ページとして作成されないため、DocuWorksViewer では取り込んだ文書の内容は表示されません。

(検索対象となる範囲)

検索エンジンで DocuWorks ファイルのインデックスを作成する際は、ページデータに含まれているテキスト情報のみが抽出され、インデックスが作成されます。

作成時に一緒に取り込んだオリジナル文書はインデックス作成の対象となりません。また、作成時に取り込んだものが画像だった場合、表示上は文字がありますが実際のテキスト情報は含まれないため、その部分のインデックスは作成されません。

図示すると下図の赤枠で囲われた部分のインデックスが作成され、検索対象となります。

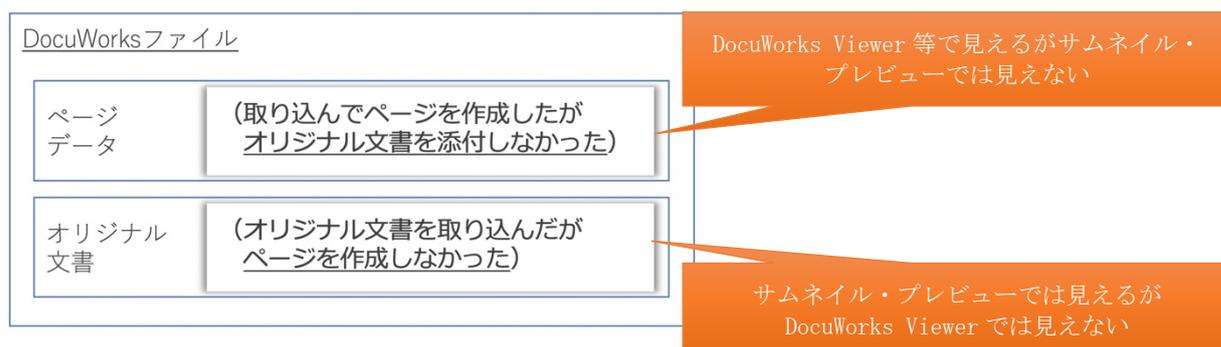


(サムネイル・プレビュー対象となる範囲)

検索エンジンで検索した際に、概要やプレビューをクリックすると、該当ファイルのサムネイルやプレビュー (PDF) が表示されます。DocuWorks ファイルもサムネイル・プレビューの対象ですが、検索エンジンの仕様上、サムネイル・プレビューが生成される範囲は、DocuWorks ファイルに取り込まれたオリジナル文書が対象となります。



このため、DocuWorks Viewer で開いてみた場合と、検索エンジンで表示されるサムネイル・プレビューの内容が異なる場合があることにご注意ください。特に以下のような構成のファイルの場合、別なファイルと認識してしまう可能性があります。



第2章 基本的な使い方

第2章. 基本的な使い方

2.1. 検索画面へのアクセス

ブラウザを起動し、下記 URL にアクセスすると検索画面が開きます。

```
http://Neuron サーバー/ui/
```

2.2. 検索画面について

Neuron ES は、大量のコンテンツの中から必要な文書をすばやく探し出す、オープンソースを活用した企業向け全文検索システムです。

インターネット検索のように Web 画面から検索キーワードを入力し、検索ボタンをクリックするだけで簡単に目的の情報を検索することができます。

Neuron ES の画面は 2 つの画面と 3 つの表示領域から構成されています。

The screenshot displays the Neuron ES search interface. The top navigation bar includes the Neuron logo and search input fields. The left sidebar contains filters for folders, tags, and file sizes. The main content area shows search results for 'neuron', including links to company pages and case studies. The bottom section features a diagram illustrating the search process: 'データ収集' (Data Collection) leads to 'データ蒸精' (Data Refinement), which then feeds into '自動分類' (Automatic Classification), '自動分析' (Automatic Analysis), and finally '異常検知' (Anomaly Detection). The diagram is titled 'データ収集から異常検知までトータルサポート' (Total support from data collection to anomaly detection).

第2章 基本的な使い方

2.2.1. 検索入力部

検索入力部はその名の通り検索キーワードを入力したり、条件を指定したりする部分になります。



(1) 検索キーワード入力欄



入力欄に検索したいキーワードを入力し、リターンキーを押下するか虫眼鏡のマークをクリックすることで検索することができます。

このキーワードは文章・単語どちらでもかまいません。Neuron ES は形態素解析+2 文字分割(2-Gram)で文章のインデックスを作成しており、検索キーワードも同様に形態素解析+2 文字分割(2-Gram)で解析された後にマッチする文書が検索されます。そのため、ある一部分しかマッチしていない文書も検索結果に表示されます。

また、入力された全角英数字は半角英数字に変換されて検索されます。入力された半角カナは全角カナに変換されて検索されます。インデックス作成時と同じ変換が行われています。

形態素解析時に抽出された語が動詞の場合は基本形に変換され、その語でインデクシング・検索されるため、完全に同一の言葉でなくても検索にヒットします。

完全に入力したキーワードにマッチした文書を表示させたい場合は、”(ダブルクォーテーション)でキーワードを括るか、完全一致チェックボックスを利用ください。

例) ”ニューロン”

AND 検索を行いたい場合はスペース区切りでキーワードを列挙ください。

例) Neuron ブレイنزテクノロジー

OR 検索を行いたい場合はキーワードとキーワードの間に太文字で“OR”と入力ください。

例) Neuron OR Impulse

含んでほしくないキーワードがある場合は、そのキーワードの前に -(ハイフン)をつけます。こうすることで、そのキーワードが含まれない検索結果を表示することができます。

例) -論文

第2章 基本的な使い方

記号(#, \$, &, @, ☆, ...等)は除去されるため、検索キーワードとして使用することはできません。

基本的に名詞・動詞・形容詞が検索キーワードとして有効です。副詞、接続詞等は検索キーワードとして利用できないとお考えください。

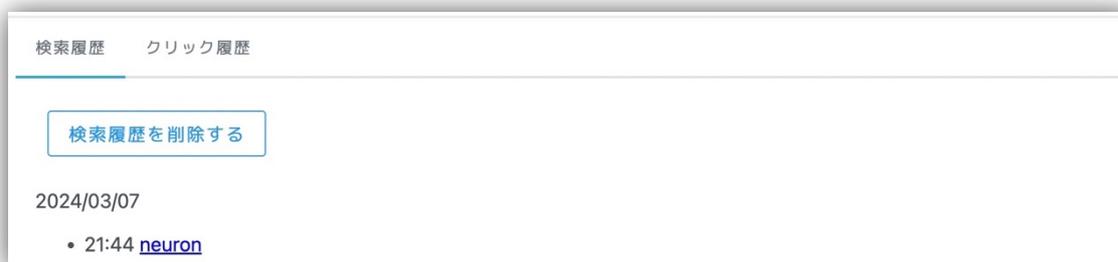
* を使ったワイルドカード検索は英数字のみ有効です。

例) neur*, 090* . . . OK
ニューロ* . . . NG

(2) 検索履歴



検索履歴アイコン(上図青丸)をクリックすると、これまで検索したキーワードの履歴と、クリックしたコンテンツの履歴を最大で30日分確認することができるページが新たに表示されます。この履歴はブラウザ(ユーザー)毎の設定であり、全ユーザーに共有される履歴ではありません。ブラウザのローカルストレージ設定が有効になっている場合に利用することができます。ブラウザを変更したり、ブラウザのキャッシュをクリアするなどした場合、これまでの履歴が削除され、見ることができなくなります。



(3) 検索対象指定



検索対象指定には、「内容+ファイル名」「フォルダ名」「ファイル名」のチェックボックスがあります。

内容+ファイル名 :ファイル名を含む全文が検索対象になります
 フォルダ名 :フォルダ名のみ検索対象になります
 ファイル名 :ファイル名のみ検索対象になります

上記から検索対象を選択することができます。フォルダ名はファイルの保存先しか覚えていないような場合に使用すると便利です。

第2章 基本的な使い方

(4) ソート



表示順リストにて、結果の並び順を変えることができます。並び順は以下の7つから選択することができます。

並び順	説明
関連度順	BM25 というアルゴリズムでスコア計算されたスコア順に表示されます。基本的に <u>キーワードの出現頻度と文書の長さ</u> が影響しているとお考えください。 短い文書にキーワードが多く含まれていればスコアが高くなります。
新しい順	最終更新日付の新しい順に表示されます。スコアは一切関係がありません。
古い順	上記とは逆に最終更新日付の古い順に表示されます。
人気順	Neuron ES を利用している全ユーザーのクリック履歴から多くクリックされた文書が上位にくるように表示されます。
いいね順	検索結果のいいねボタンが多くクリックされた文書が上位にくるように表示されます。
ファイル名(昇順)	半角→全角の順で並び替えを行います。それぞれの中で記号→数字→アルファベット→(大文字→小文字)→日本語の順で表示されます。具体的には UTF-8 の文字コード順になります。
ファイル名(降順)	上記昇順の逆の順序で表示されます。

既定は関連度順で表示されます。

(5) オプション指定



完全一致・類義語を使用することができます。

完全一致を ON にした場合、原則キーワードに完全に一致する検索結果が表示されます。ただし、完全一致といえども「ブレインズの Neuron ES」といったキーワードの場合、処理の仕様上「ブレインズの Neuron ES」ではなく「ブレインズ Neuron ES」といったように「の」が省いたもので検索されます。

また、キーワードを完全に含む語もヒットします。

例) 「ブレインズ」で検索

下記がヒット

- ブレインズ
- ブレインズテクノロジー
- テクノブレインズ

類義語機能はシステム管理者側で登録した類義語に従って検索が行われます。標準では、表記ゆれが登録されています。表記ゆれとは「打ち合わせ、打ち合せ、打合せ、打合わせ」などと同じ意味を持つ言葉について表記が混在しているものを指します。

それ以外の類義語については登録がある場合のみ機能します。

第2章 基本的な使い方

(6) 詳細検索

含む

内容 + ファイル名

AND

内容 + ファイル名

内容 + ファイル名

内容

フォルダ名

ファイル名

タイトル

追加

追加

上記の内容で検索する

完全一致・類義語チェックボックス横の詳細検索をクリックすると上記のような画面が表示され、より複雑な検索を行うことができます。追加ボタンをクリックすると入力欄が増え、どの対象に対して検索したいか、AND 検索・OR 検索を行いたいかを指定できます。また、含む検索なのか、含まない検索なのかも指定することができます。

(7) 個人設定

ログインID

mirai.taro@brains-tech.co.jp

個人設定

ページあたりの表示件数

10件 20件 30件 50件 100件

検索キーワードマッチ部 ハイライト

太文・斜字 色付き背景

個人設定をクリックすると上記のような画面が表示され、検索結果の表示件数・ヒットした箇所のハイライト方法を変更することができます。

第 2 章 基本的な使い方

2.2.2. ファセット部

Neuron ES では以下の 7 種類で検索結果を絞り込んでいくことができます。

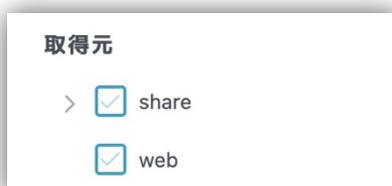
(1) フォルダ指定



フォルダ指定はエクスプローラーのようにフォルダを指定して絞り込みをすることができます。なお、クリックしたフォルダがサブフォルダを含んでいる場合、そのサブフォルダも含んだ絞り込みとなります。該当フォルダのみの絞り込みはできません。また、この表示はサーバー名も含めて 6 階層目まで、かつアクセス権があるフォルダしか表示されません。

何も指定しない場合、アクセス権を持つすべてのフォルダから検索されます。

(2) 取得元



大きく Web 系とファイルサーバー系で登録されており、さらに各カテゴリでサブカテゴリが登録されています。選択したもので絞り込んだ結果が表示されます。何も選択しない場合、両方から検索した結果が表示されます。(お客さま設定によってこの項目の数・名称は変更されます)

(3) ファイル更新日



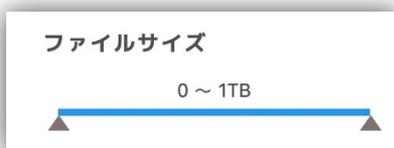
指定なし、2 日以内、1 週間以内、1 ヶ月以内、1 年以内、3 年以内、および期間指定による絞り込みができます。

第2章 基本的な使い方

期間指定で絞り込む場合、YYYYMMDD形式で入力する必要があります。

例) 2021年1月1日から2021年6月30日までのファイルを探す場合
20210101 - 20210630

(4) ファイルサイズ



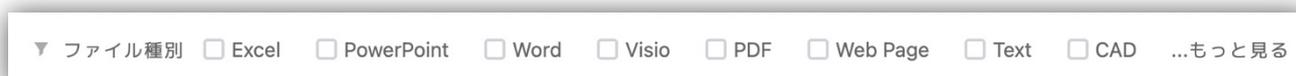
スライダーを動かすことで範囲指定での絞り込みができます。

(5) 最終更新者



Officeファイルの最終更新者で絞り込みができます。最終更新者がかな・カタカナ・漢字の場合は部分一致、英数字の場合は完全一致での絞り込みになります。Officeファイル以外には使用できません。

(6) ファイル種別



OfficeファイルやPDF等の文書の種類で絞り込むことができます。

なお、ファイル種別が表示される種類は、インデックス作成したファイルに含まれるファイルの種類に依存します。何も選択しない場合、すべてのファイル種別から検索した結果が表示されます。

(7) タグ検索

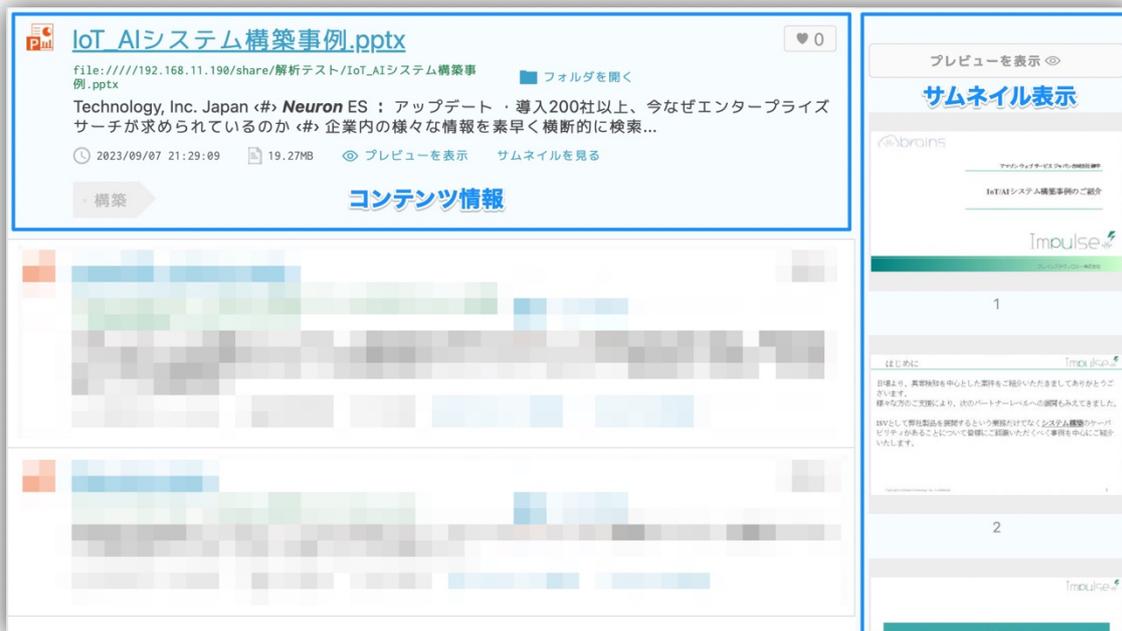


インデックス作成時に付与されたタグで絞り込むことができます。

なお、タグの数が多い場合は右側のタグ検索欄にタグ名を入力することでタグ探して絞り込むことができます。このタグ検索は部分一致で検索されます。

第 2 章 基本的な使い方

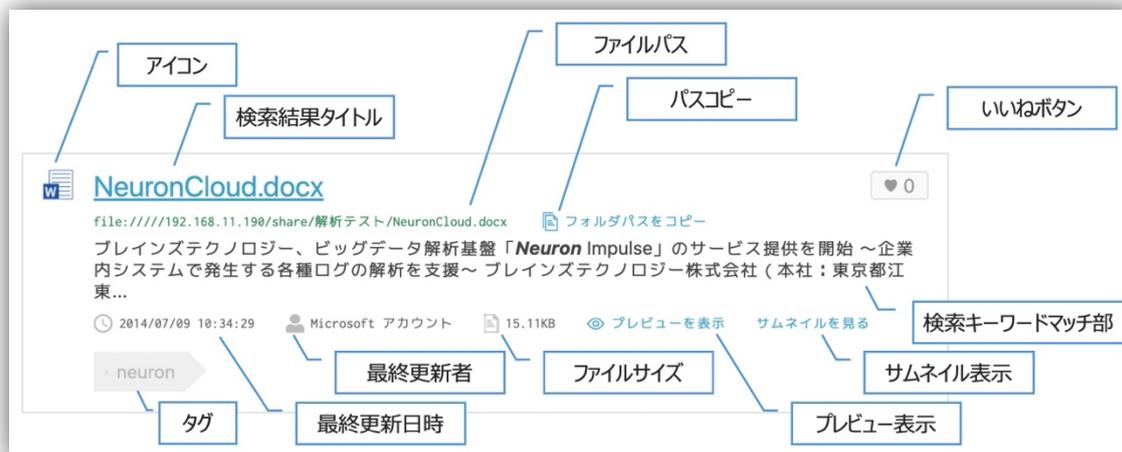
2.2.3. 検索結果表示部



結果表示部では、検索キーワードに適合し、かつそのコンテンツにアクセス権がある

- コンテンツ情報
 - サムネイル表示(対象ファイル時)
- が表示されます。

(1) コンテンツ情報



情報	説明
アイコン	ファイル種別に応じたアイコンが表示されます。
検索結果タイトル	Web ページの場合はタイトル、それ以外の場合はファイル名が表示されます。Office や PDF 等の場合、クリックするとそのコンテンツを開くことができます。検索結果がファイルサーバーの場合で、かつ Edge ブラウザの IE モードで閲覧している場合は直接そのコンテンツを開き、そうでない場合は該当ファイルをダウンロードします。
ファイルパス	ファイルの実際の場所情報が表示されます。
パスコピー	そのファイルが保存されているパス情報をクリップボードにコピーします。なお、検索結果がファイルサーバーの場合で、かつ Edge ブラウザの IE モードで閲覧

第2章 基本的な使い方

	している場合、もしくは openapp というツールを導入し管理者が表示するよう設定している場合は「フォルダを開く」が表示されます。
いいねボタン	検索結果に対して「いいね」をつけることができます。
検索キーワード マッチ部	キーワードがヒットした前後の文が表示され、ヒットした部分(ハイライト)は太字・斜体文字で表示されます。なお、ヒットした箇所がファイル名の場合は、先頭から 100 文字抽出されたものが表示されます。
最終更新日時	ファイルの最終更新日時が表示されます。
最終更新者	Office ファイルの最終更新者が表示されます。なお、設定されていない場合は Office ファイル以外は表示されません。
ファイルサイズ	ファイルのサイズが表示されます。
プレビュー表示	Office ファイルを PDF 化して別タブで表示します。なお、その際、検索キーワードはハイライトされて表示されます。サムネイルと同様にクリック時に生成されるため、表示までに時間がかかります。ページ数が多いコンテンツの場合、数十秒かかる場合があります。
サムネイル表示	Office ファイルや PDF 等を JPEG 画像化して表示します。クリック時に生成されるため、表示までに時間がかかります。
タグ	インデックス作成時にタグが付与されている場合に表示されます。

(2) サムネイル表示

サムネイル表示対象である場合、表紙を含め最大 4 ページ分のサムネイルが表示されます。なお、この表示で使用されるサムネイルは、サーバー上にキャッシュがなければ概要リンクがクリックされた際に生成されるため、表示までに時間がかかります。

また、表示されたサムネイルの上にある「プレビューを表示」ボタンをクリックすると別タブが開き、PDF 化された文書が表示されます。

なお、サムネイル表示対象は Microsoft Office 形式のファイル、PDF、画像ファイル、DocuWorks に限定されます。それ以外は表示されません。

第2章 基本的な使い方

2.2.4. プレビュー表示部

検索結果のプレビューをクリックしたときに開かれる画面になります。



検索キーワードが1単語の場合のみヒットした箇所が薄緑色、または薄赤色でハイライトされます。薄緑色は現在ハイライトされている箇所、薄赤色は候補のハイライトになります。検索で \wedge / \vee をクリックすることで薄緑色のハイライトが移っていきます。

(1) 縮小版表示

PDF 化された各ページの縮小画像が表示されます。



(2) 検索メニュー

PDF 内を検索するためのツール類が表示されます。



現在の検索単語は、検索画面に入力したものが引き継がれます。ただし、複数キーワード入力時はキーワード入力欄が赤くなり、ヒット箇所を表示できなくなることに注意ください。

(3) ズームメニュー

PDF 表示の倍率を変更することができます。デフォルトでは自動ズームになっています。「+」をクリックすると倍率が上がり、「-」をクリックすると倍率が下がります。自動ズームと表示されている部分ををクリックするとリストから倍率を選択することができます。

第 2 章 基本的な使い方

(4) 印刷等メニュー

表示されている PDF を印刷、保存、または別な PDF を開くメニューです。

(5) 編集メニュー

表示されている PDF に注釈を入れたり、あるいはペンツールで書き込みをすることができます。
文字の色、ペンの色を変更することも可能です。

第 2 章 基本的な使い方
