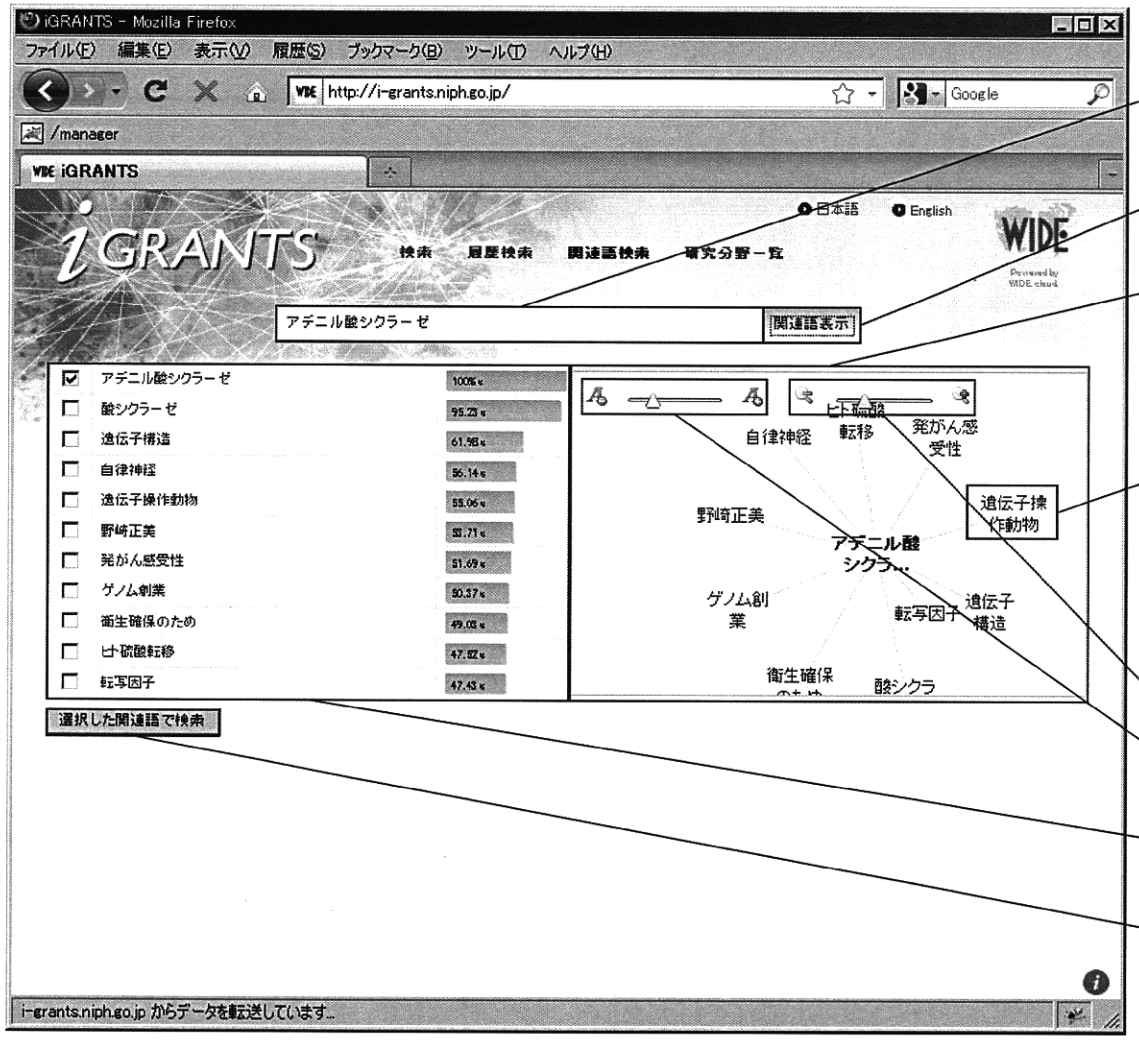


キーワード入力テキストフィールド。
キーワードの断片を入力するとiGrantsにキーワードとして登録しているものの中から候補がサジェスト表示される。ユーザはその候補からキーワードを選択する。

入力されたキーワードと類似性の高いキーワードを表示する。ただし、iGrantsにキーワードとして登録されていない文字列が入力された場合は表示しない。

ユーザによって入力されたキーワードと類似性の高いキーワードをバネグラフ表示する。

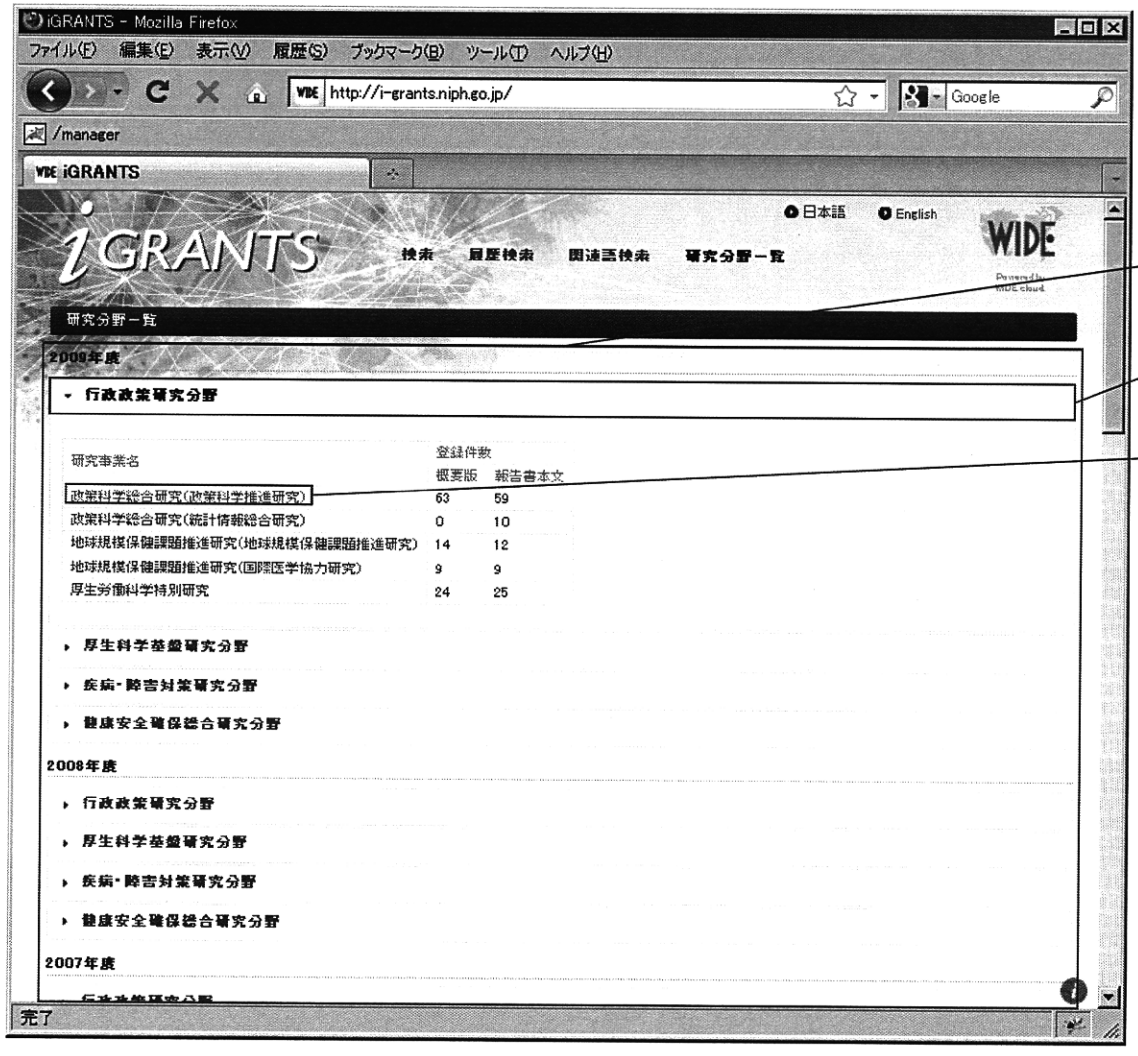
ノード。
表示されている文字列はキーワードを表わしている。
文字列長の長いキーワードについては省略表示する。
ダブルクリックするとそのキーワードで報告書を全文検索する。また、右クリックで以下のメニューが選択可能。
「関連語を表示」…そのノードを中心に類似性の高いキーワードをバネグラフに追加する
「項目を削除」…そのノードをバネグラフから削除する。削除によって孤立するノードについても削除する。

ノード間の斥力を調節可能

バネグラフの各ノードの文字の大きさを調節可能

ユーザによって入力されたキーワードと類似性の高いキーワードをその類似度と共に棒グラフで表示する。

上で選択されたキーワード間で報告書のOR検索を行なう。



年度＞分野＞事業の順で研究分野が階層表示される。

クリックすると該当する分野に属する事業の表示／非表示を切り替えることができる。

クリックすると該当する事業名で報告書の全文検索を行なう。

検索結果の表示形式を設定することができる。
・1ページあたり何件ずつ表示するか？
・何順で表示するのか？

ページNoへのリンク。最大10ページ分を表示。

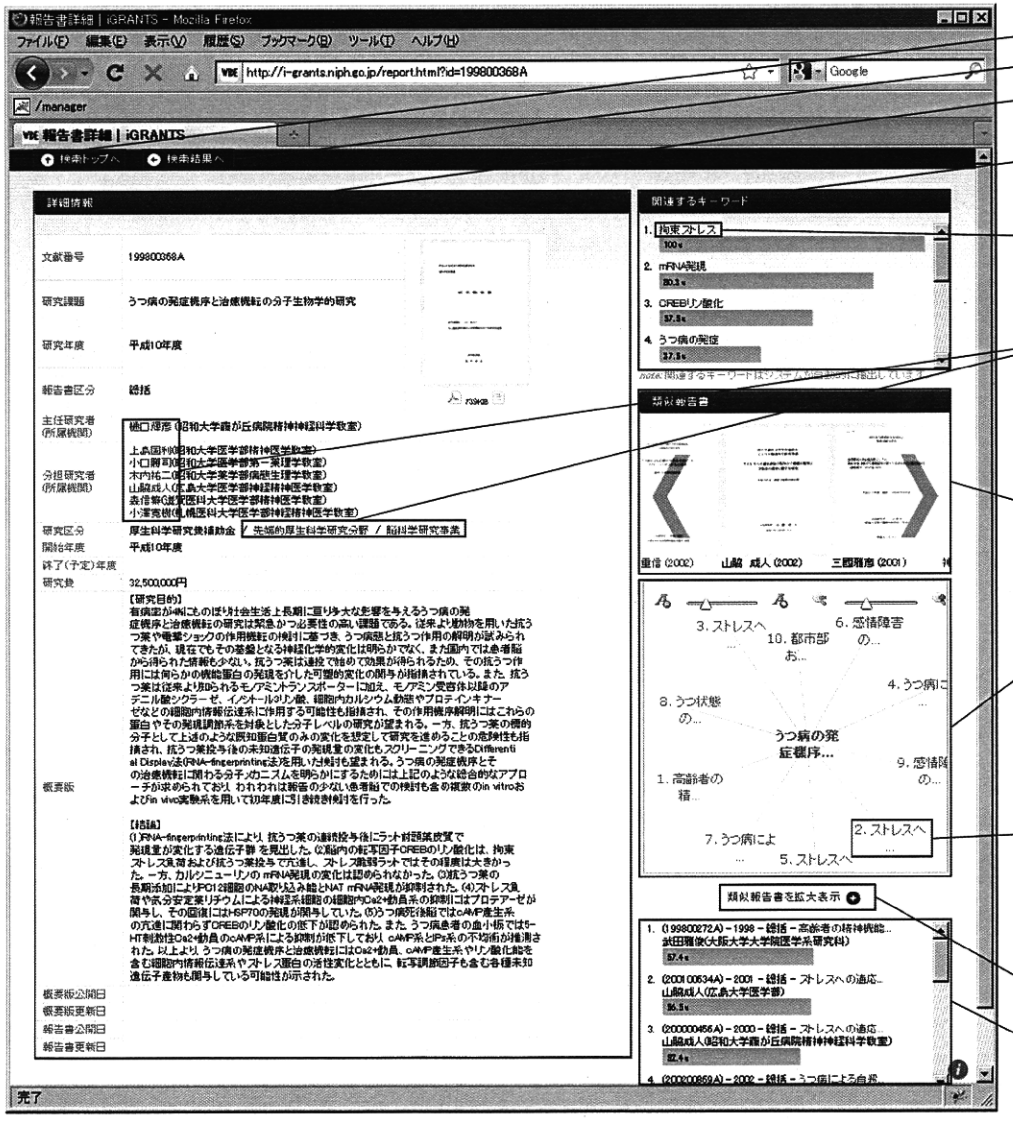
表示中にページに対して、「次」のページおよび「前」のページへのリンク

検索によりヒットした報告書のサムネイル画像。

検索によりヒットした報告書の概要。
左から、
・文献番号
・研究年度
・報告書区分
・研究課題
となっている。

検索によりヒットした報告書の詳細情報。
・主任研究者
・研究区分
・(関連する)キーワード
は常に表示される。それ以外の項目(本文など)は、検索によりヒットした項目がスニペット表示される。

報告書をPDFファイルまたはテキストファイルでダウンロードするためのリンク。



トップページへのリンク

検索結果ページへ戻る為のリンク

報告書の詳細情報

表示中の報告書が持つ特徴的なキーワード。
重要度順に最大10件まで棒グラフ表示。

クリックするとそのキーワードで全文検索

クリックするとその研究者名や分野名で全文検索

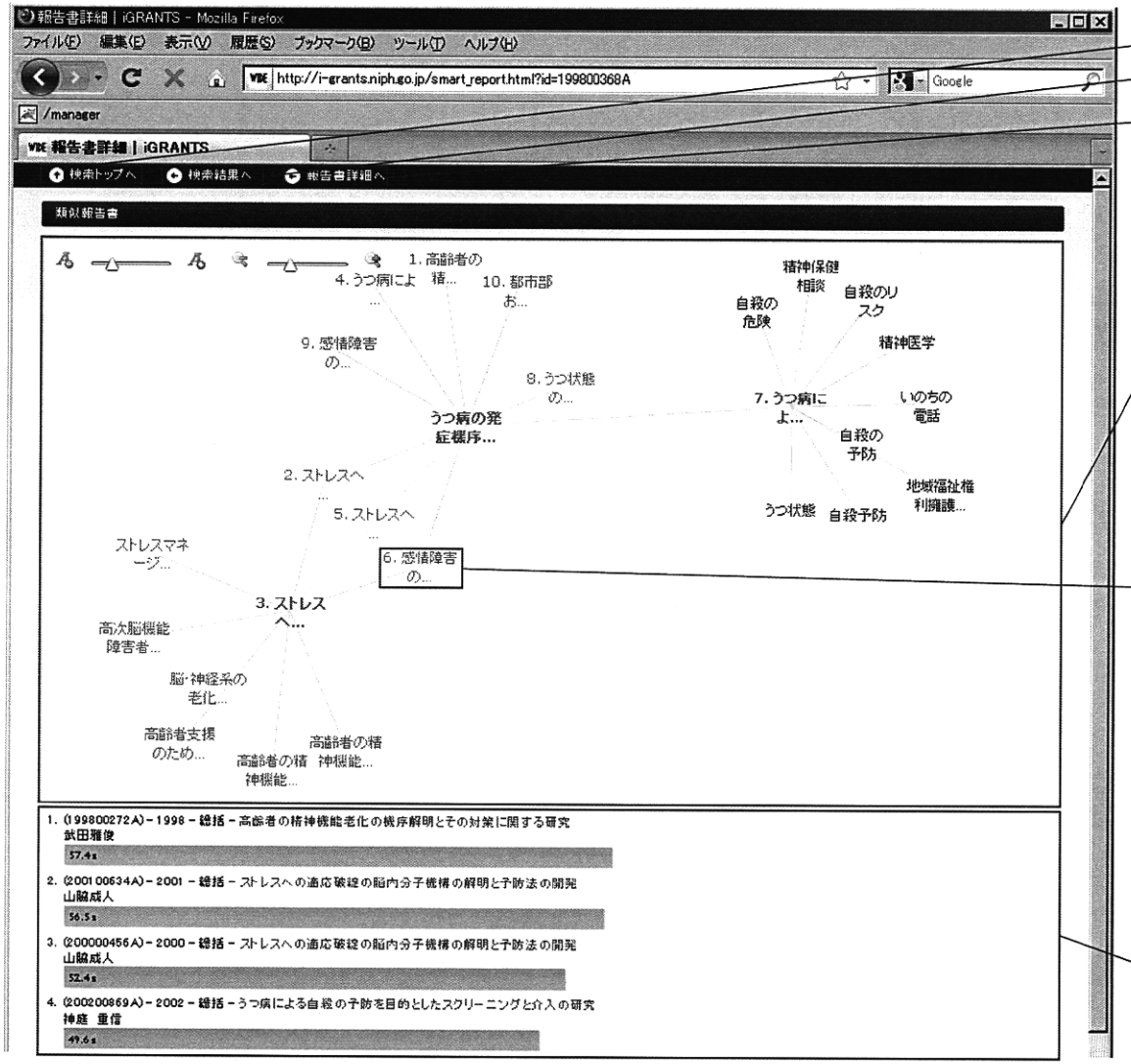
表示中の報告書に類似している報告書のサムネイルをギャラリー表示。
サムネイルをクリックすることで、その報告書ページへジャンプする。
「<」や「>」の上にマウスポインタを合わせると、左や右へスクロールする。

表示中の報告書に類似している報告書のサムネイルをバネグラフ表示。

ノード。
表示されている文字列は研究課題を表わしている。文字列長の長い研究課題については省略表示する。ダブルクリックするとその報告書ページへジャンプする。
また、右クリックで以下のメニューが選択可能
「関連語を表示」…その報告書ノードを中心に特徴的なキーワードノードをバネグラフに追加する
「類似報告書を表示」…その報告書ノードを中心に類似性の高い報告書ノードをバネグラフに追加する
「項目を削除」…そのノードをバネグラフから削除する。削除によって孤立するノードについても削除する。

類似報告書ページへのリンク

表示中の報告書に類似している報告書のサムネイルを棒グラフ表示。



トップページへのリンク

検索結果ページへ戻る為のリンク

報告書ページへ戻る為のリンク

表示中の報告書に類似している報告書のサムネイルをバネグラフ表示。

赤ノードは報告書ノード。
緑ノードはキーワードノード。

文字列長の長いノードについては省略表示。
赤ノードをダブルクリックするとその報告書ページへジャンプする。緑ノードをダブルクリックすると、そのキーワードで報告書を全文検索。

また、右クリックで以下のメニューが選択可能
「関連語を表示」…その報告書ノードを中心に特徴的なキーワードノードをバネグラフに追加する
「類似報告書を表示」…その報告書ノードを中心に類似性の高い報告書ノードをバネグラフに追加する（赤ノードのみ）
「項目を削除」…そのノードをバネグラフから削除する。削除によって孤立するノードについても削除する。

表示中の報告書に類似している報告書のサムネイルを棒グラフ表示。

III-4. 研究者メールアドレスの復元手法

研究者メールアドレスの復元手法

はじめに

過去に紙媒体で提出された研究報告書について、改めてPDF として各研究者に提出を要請するためには、過去の研究課題の研究者に対して、emailにてその旨を伝える必要がある。しかしながら、GRANTSにおける研究課題の一覧が収められたデータベース(以下「課題データ」と記す)には、各課題の研究者名は含まれているものの、emailアドレスは含まれていない。そこで、「課題データ」とは別に、研究者の一覧が収められたデータベース（以下「研究者データ」と記す）を参照し、そこに収められている研究者名とemail アドレスを利用する必要がある。本稿では、この2つのデータベースのレコードをお互いに対応付けるための手法について記す。

事前処理

「課題データ」の各レコードには、研究課題、研究者名と組織名、研究開始年、研究費などのデータが含まれ、レコードの数は8,078件である。「課題データ」の例を図1に示す。「研究者データ」の各レコードには、研究者名、email アドレス、研究者ID、機関名、卒業大学名、卒業年、学位などのデータが含まれ、レコードの数は13,365件である。「研究者データ」の例を図2に示す。「課題データ」の各レコードの研究者を、「研究者データ」内から自動的に探して対応付けるのが、今回の操作の目的である。

2つのデータベースに共通して含まれているデータの種類として、研究者名と組織名が挙げられる。特に、研究者名が一致した2つのレコードについては、その確度も高いと考えられる。そこで、まず研究者名が一致するレコードのペアの一覧を得ることを考える。「課題データ」上では、研究者名と組織名が単一フィールドにまとめて入力されているため、まずそれらを分離する操作を行う。また、各データベースの研究者名には、表記の揺れ(姓と名の間のスペースの有無など)があり、そのままでは比較の際に扱いづらい。そこで、表記を統一するため、すべての研究者名の姓と名は、間を開けずに続けて記すようにした。

姓名に基づく名寄せ操作

この結果得られた研究者名について、2つのデータベースの各レコードが一致するか試みたところ、5,624件の対応するレコード対を取得した。なお、今回は、研究者の姓名が一致しない2つのデータベースのレコードについては、別人であるとみなして操作を行なった。しかし、登録の時期の違いによって、姓を変えた同一人物が2つのデータベースにそれぞれ異なる研究者名で入力されているという場合も考えられる。そうした場合においても、同一人物を対応付けたいのであれば、「研究者データ」に男性として記載のある者以外の研究者を対象として、名のみによって同一人物の候補を探し、後述する名前以外の類似度の算出によって候補を絞り込むことによる候補取得の方法が考えられる。

しかし、こうした操作によっても、名が一致しないレコード対は、姓名が一致している対に比べて、同一人物である確からしさが低いと考えられる。職歴などにたまたま共通点があり、名が同じであれば、同一人物とみなされかねないためである。そのため今回は、姓名が一致したレコード対のみを同一人物である候補とみなすことにする。また、「研究者データ」の研究者名に旧名が付記されているレコードが24件ほど確認できた。これらのレコードについては、まれな例であるため自動的な比較の対象としていない。

組織名に基づく名寄せ操作

先の操作で得られたレコード対は、姓名の一致のみを元に判断しているため、同姓同名の別人を対応付けていることも考えられる。そこで、組織名の一致によって、得られた対応が正しいか判断することを考える。組織名については表記が統一されておらず、研究者名の場合のように完全一致で判断すると、ほとんどが一致しない対であるという結果になるため、類似度を算出することで対応する。2つの組織名の類似度の算出のためには、まずそれぞれの組織名を2gramの集合で表す(例: 厚生労働省 ⇒ (厚生、生労、労働、働省))。そして、こうして表された、2つの組織名の積集合の大きさを、大きさが小さい方の組織名の大きさで割った値を、類似度として用いた。

組織名の類似度を測る方法は有用であるが、この方法では組織を移るなどして2つのデータベースの対応するレコードの組織名が異なっている場合については対処できない。そこで、2つのレコードの対応が同姓同名の他人のものではなく、同一人物の対応であることを確かなものとするために、個人の情報を提供するWebサイトを利用することを考えた。図3に概要を示す。今回は、利用するWebサイトとして、JGLOBAL(http://jglobal.jst.go.jp/)とSPYSEE (http://spysee.jp/)を対象とした。利用するWebサイトの各ページには、姓名、現職、職歴、卒業大学名、卒業年などの個人情報が記載されており、研究者名およびペアの各レコードに含まれる少なくとも1つのデータとマッチするページが存在すれば、ペアの対応は尤もらしいと判断できる。

結果

以上の方法によって、「課題データ」のなかから「研究者データ」に含まれる研究者を抜き出して、レコード対として対応付けることができた。適当な類似度の閾値において、対応付けられたレコード対の研究年別の件数を図4に、対応付けの方法の種類ごとの一致件数を図5に示す。これらの方法によって自動的に確度が検証されたレコード対のうち、依然としてその対応が不確かであるものは、約1,000件ほどである(この件数は閾値とする類似度の値によって変動しうる)。

課題	研究年	研究者・組織	研究費
薬物に関する研究	1999	田中 太郎 (T大学医学部教授)	12,000,000
老化に関する研究	2000	鈴木一郎(I医科大学)	47,000,000
看護に関する研究	2003	厚生太郎 <A医療センター>	800,000

図 1 課題データ

研究者ID	研究者名(漢字)	研究者名(カナ)	Email	機関名	卒業大学	卒業年
123	田中 太郎	タナカ タロウ	tnk@example.jp	T大学 医学部	T大学	1980
456	労働 花子	ロウドウ ハナコ	hnk@ex ample.jp	D製薬株式会社	K大学	1995
789	厚生 太郎	コウセイ タロウ	kt@exam.ple.jp	B研究所	C大学	2000

図 2 研究者データ

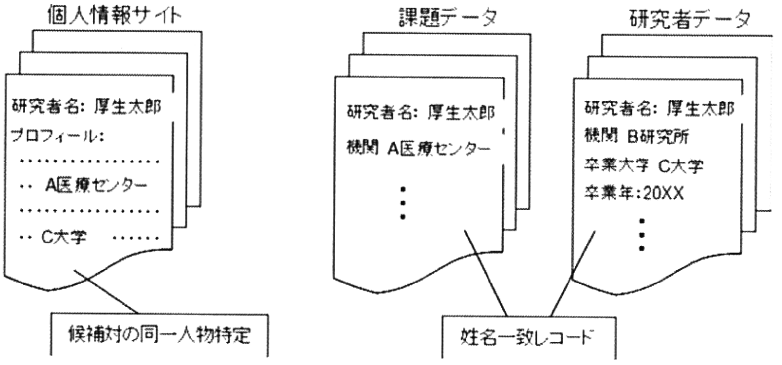


図 3 Web サイトを用いた確度の検証

研究年	高類似度	姓名一致	課題データ
1997	329 (33%)	487 (49%)	981
1998	354 (39%)	493 (54%)	901
1999	427 (45%)	580 (61%)	946
2000	565 (49%)	751 (65%)	1152
2001	733 (58%)	921 (73%)	1251
2002	959 (67%)	1147 (80%)	1430
2003	1091 (76%)	1245 (87%)	1417
合計	4458 (55%)	5624 (69%)	8078

図 4 課題データ中の年別レコード対応件数

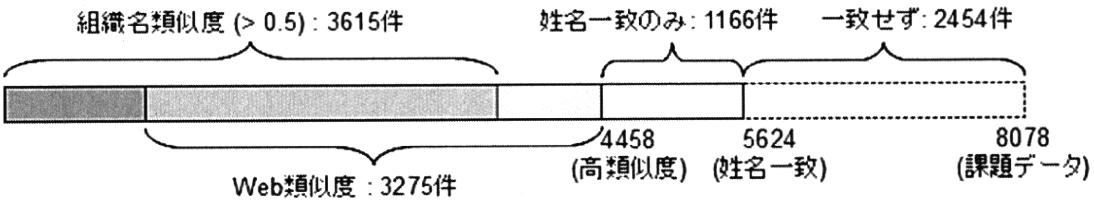


図 5 課題データ中の一致したレコード件数

III-5. 研究報告書ファイル回収用 PDF 作成マニュアル

厚生労働科学研究報告書PDF(透明テキスト付) アップロードについて

「厚生労働科学研究成果データベース」は、次年度のリプレイスを目途といたしまして国民による研究成果の一層の活用を目的に、「厚生労働科学研究成果データベース」の高機能化を進めております。なかでも「全文検索」機能は最優先で実装する予定です。さらに、公開の迅速化などを踏まえ、研究報告書をファイル形式でご提出いただくよう検討を進めている段階です。

つきましては、諸事多難な折大変お手数をお掛けいたしますが、下記報告書のPDFファイルの送付をご依頼申し上げます。ご協力いただければ幸いです。

1. アップロードファイル

＊研究年度の最終報告書のPDF(MS Word, 一太郎等より直接生成したもの)

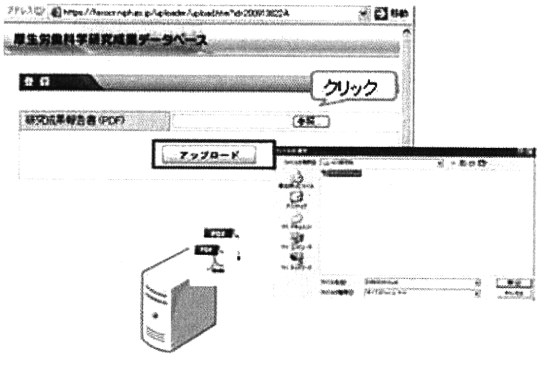
※ PDFデータを生成する前のWord、一太郎等のファイルがあれば、併せてアップロードして下さい。

＊表紙から最終ページまで、一続きのPDFファイルになるように作成して下さい。

注意: スキャニングで作成した画像PDFは対象外です。
PDFデータの作成例は、別紙をご参照下さい。
最終報告書に含まれる研究成果の刊行物を除いてPDFを作成して下さい。
PDF、MS Word、一太郎ファイルのみアップロード可能です。

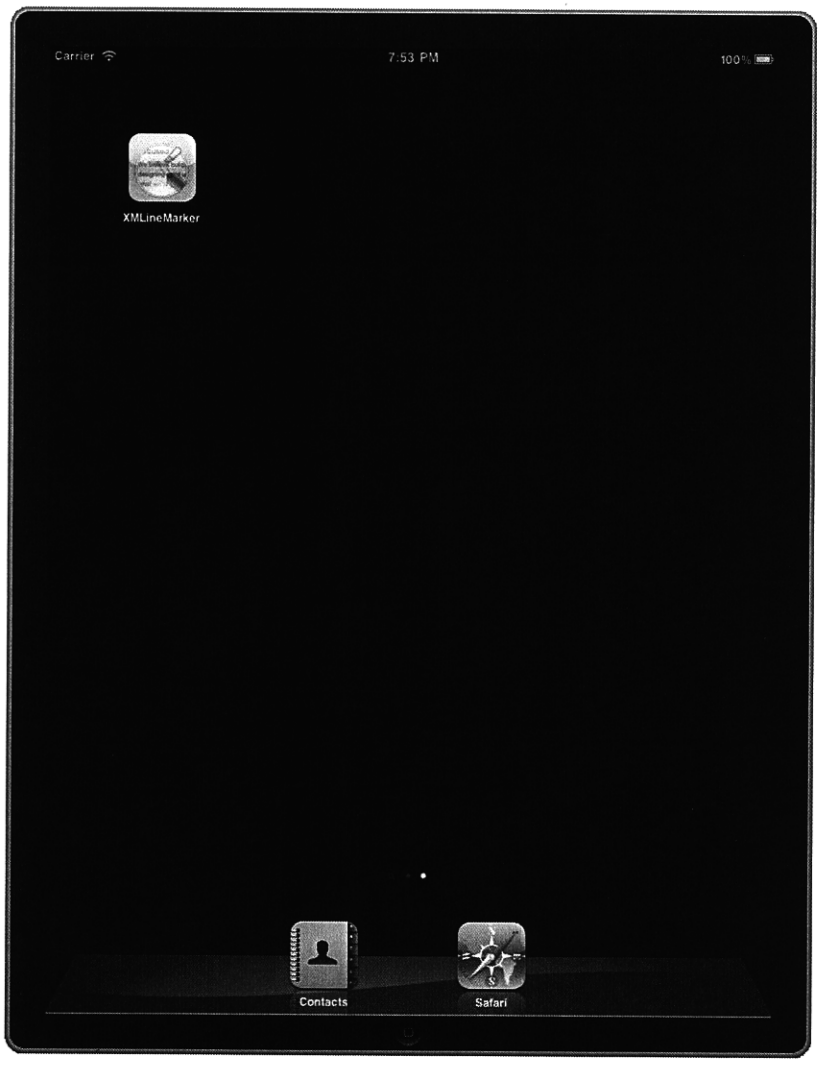
2. アップロード

1. 送信されてきた依頼メールに記述されている「アップロードURL」にアクセスする。
2. アップロードサイトの「参照」より、アップロードするファイルを選択する。
3. 「アップロード」をクリックする。

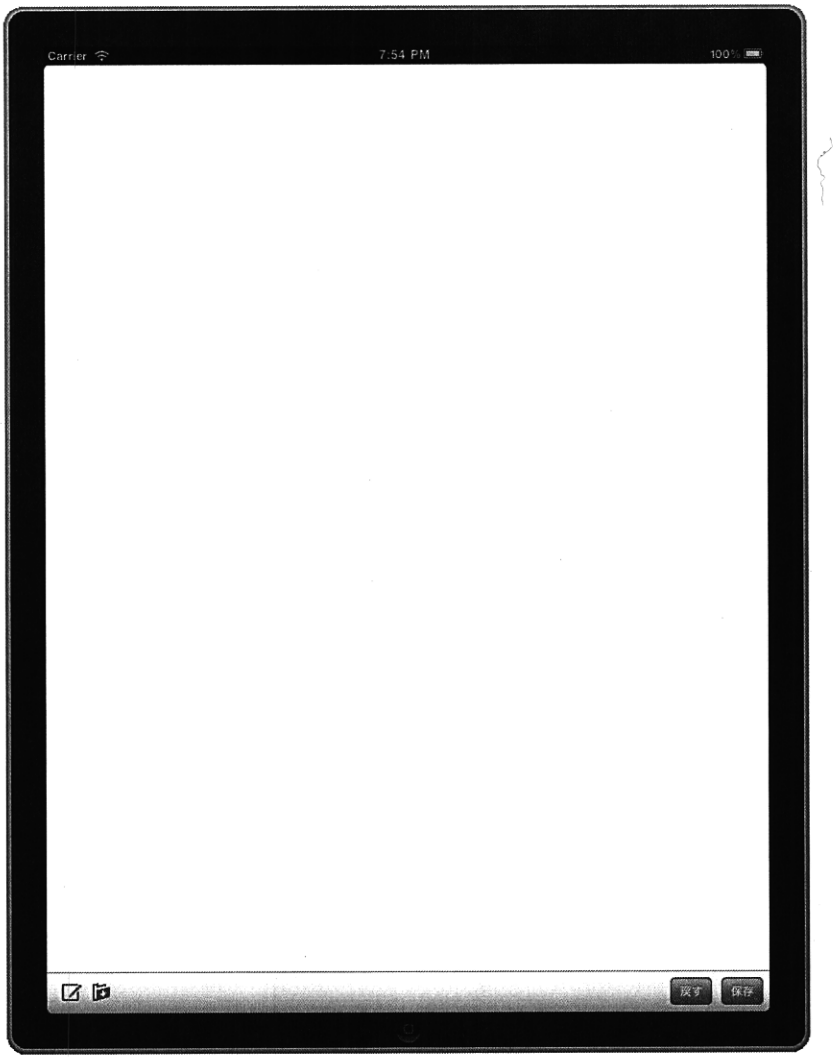


III-6. アノテーションツール操作マニュアル

アプリケーションの起動と終了



メニュー画面からアイコン(XMLineMarker)をタッチします。

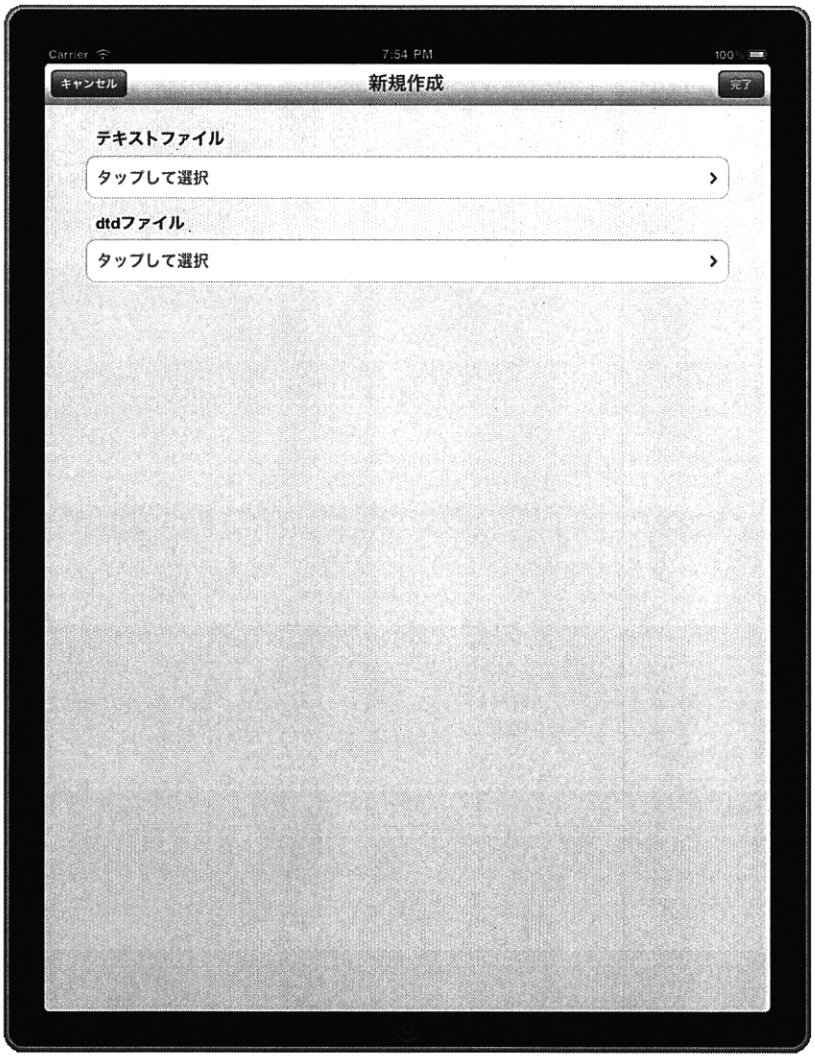


上記のホーム画面が表示されれば起動は完了です。
Homeボタン(画面下の○ボタン)を押すとアプリケーションが閉じられます。
（終了はしません。また、その他、起動終了に関する操作はiPad規定の通常の動作となります。）

XMLファイルの操作

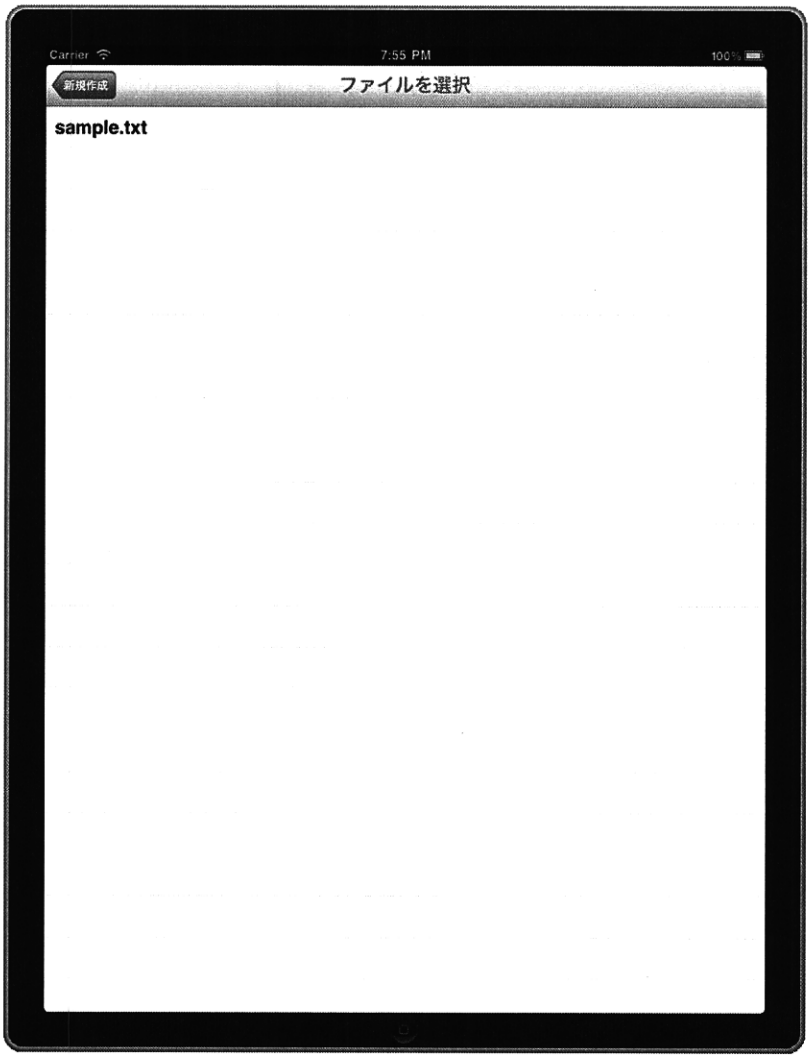
新規XMLファイルの作成

事前に用意して登録したテキストファイルDTDファイルから、編集可能なXMLを生成します。
画面左下の新規作成用のアイコンをタッチします。



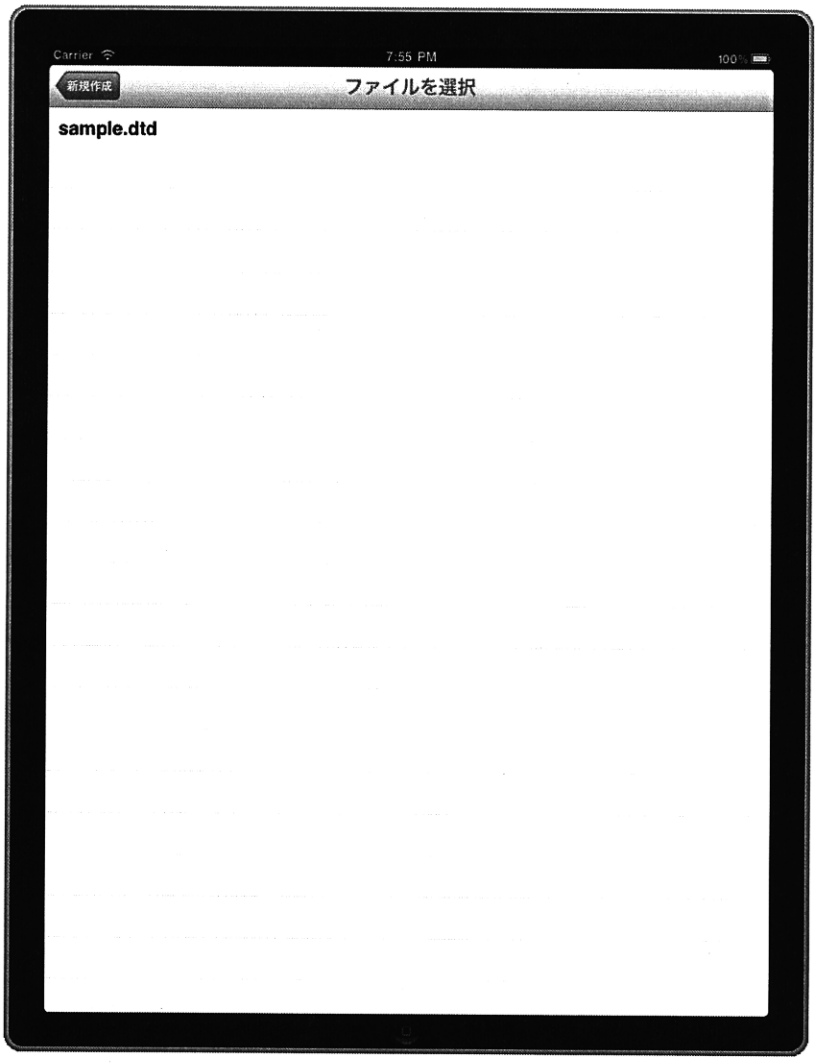
テキストファイルとDTDファイルの選択画面になります。
タップして選択からファイルを選択します。

テキストファイルの選択



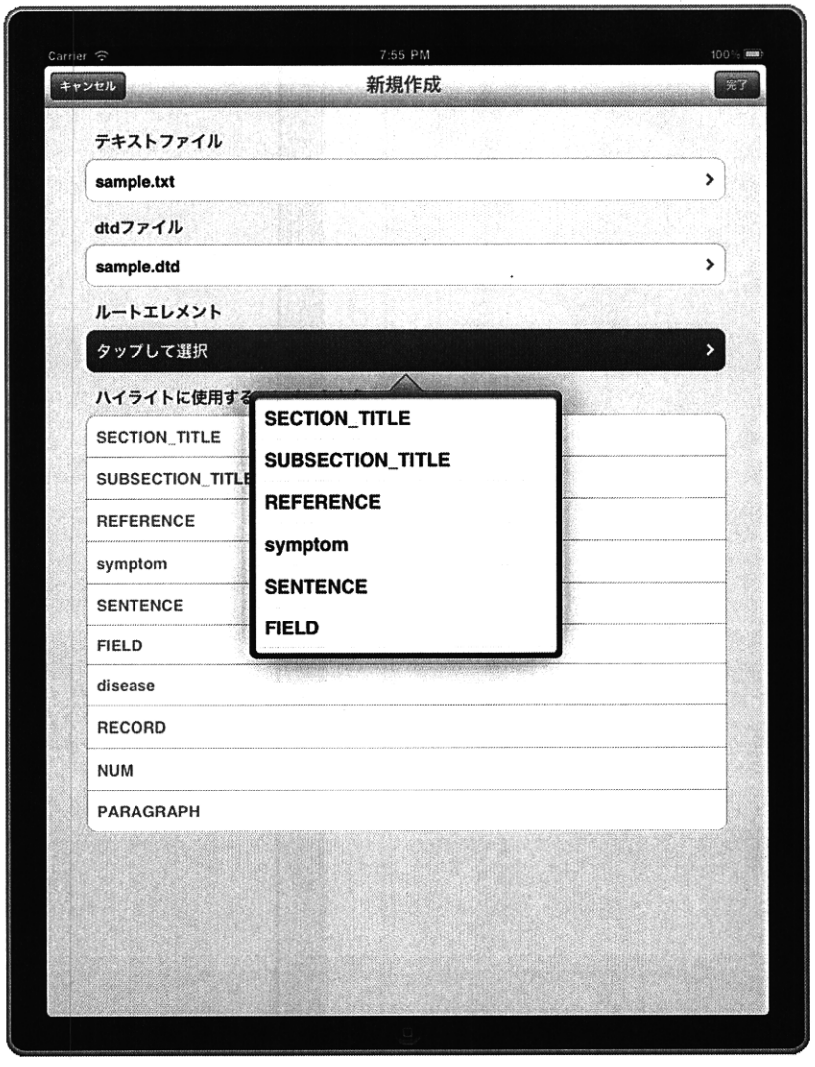
登録済みのテキストファイルが一覧表示されます。
編集したいテキストファイルを選択してください。

DTDファイルの選択



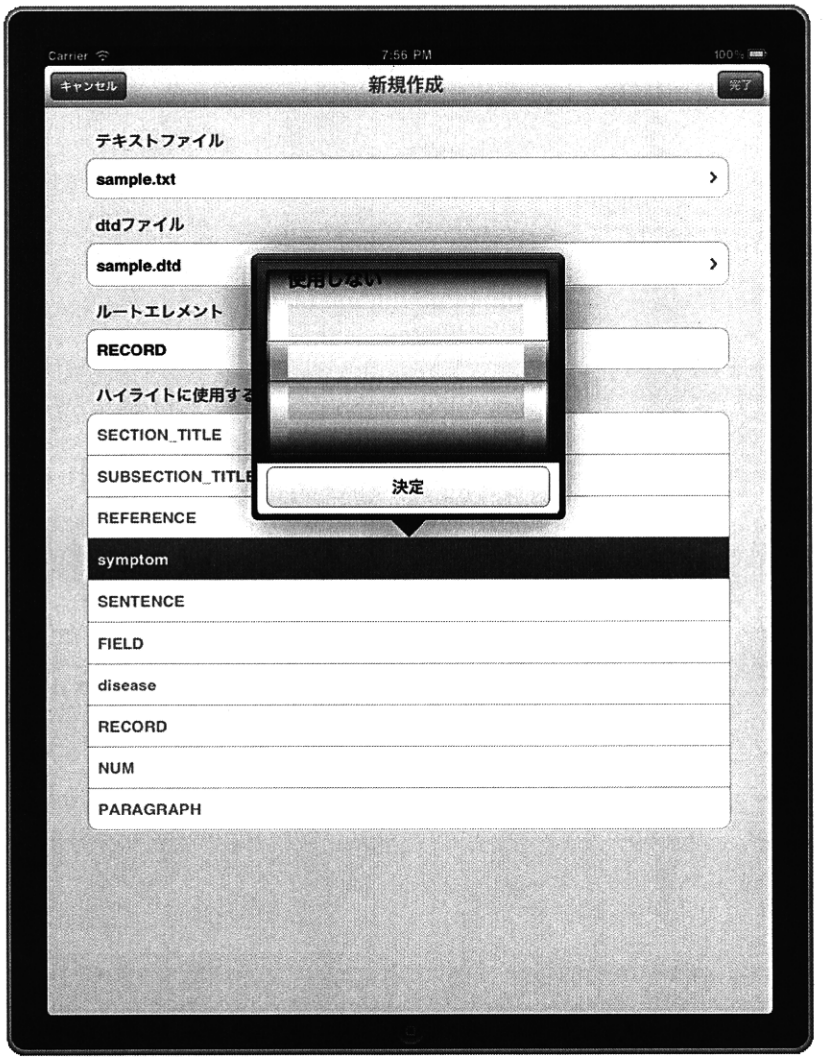
DTDファイルも同様に一覧表示されますので、適応したいDTDファイルを選択してください。

ルートエレメントの選択



DTDファイルを選択すると、ルートエレメントおよびハイライト対象のエレメント選択画面が表示されます。
編集用に作成するXMLファイルのルートエレメントを選択してください。

マーキングタグ(ハイライト色の選択)



DTDに定義されたエレメントからハイライトの対象となるエレメントを選択し、ハイライト色を決定します。