

# ManifoldCF- エンドユーザマニュアル

## Table of contents

1 概要.....	3
1.1 出力コネクションの定義.....	4
1.2 変換コネクションの定義.....	9
1.3 権限グループの定義.....	11
1.4 リポジトリコネクションの定義.....	12
1.5 通知コネクションの定義.....	18
1.6 ユーザーマッピングコネクションの定義.....	20
1.7 権限コネクションの定義.....	23
1.8 ジョブの作成.....	28
1.9 ジョブの実行.....	35
1.10 状態レポート.....	37
1.11 履歴レポート.....	42
1.12 認証についての覚え書き.....	46
2 出力コネクションタイプ.....	46
2.1 Amazon Cloud Search出力コネクション.....	46
2.2 ElasticSearch出力コネクション.....	46
2.3 ファイルシステム出力コネクション.....	48
2.4 HDFS出力コネクション.....	48
2.5 MetaCarta GTS出力コネクション.....	49
2.6 Null出力コネクション.....	49
2.7 OpenSearchServer出力コネクション.....	49
2.8 Solr出力コネクション.....	51
3 ユーザーマッピングコネクションタイプ.....	56

3.1 正規表現ユーザーマッピングコネクション.....	56
4 権限コネクションタイプ.....	57
4.1 Active Directory権限コネクション.....	57
4.2 LDAP権限コネクション.....	58
4.3 OpenText LiveLink権限コネクション.....	60
4.4 EMC Documentum権限コネクション.....	62
4.5 Memex Patriarch権限コネクション.....	65
4.6 Autonomy Meridio権限コネクション.....	67
4.7 CMIS権限コネクション.....	71
5 リポジトリコネクションタイプ.....	73
5.1 汎用ファイルシステムリポジトリコネクション.....	73
5.2 汎用RSSリポジトリコネクション.....	75
5.3 汎用Webリポジトリコネクション.....	85
5.4 Windows Share/DFSリポジトリコネクション.....	98
5.5 汎用データベースリポジトリコネクション.....	104
5.6 IBM FileNet P8リポジトリコネクション.....	110
5.7 EMC Documentumリポジトリコネクション.....	112
5.8 Dropboxリポジトリコネクション.....	115
5.9 OpenText LiveLinkリポジトリコネクション.....	118
5.10 Memex Patriarchリポジトリコネクション.....	124
5.11 Meridioリポジトリコネクション.....	126
5.12 Microsoft SharePointリポジトリコネクション.....	130
5.13 CMISリポジトリコネクション.....	136
5.14 Wikiリポジトリコネクション.....	138

## 1 概要

本マニュアルはManifoldCFを利用するエンドユーザ向けのマニュアルです。ManifoldCFフレームワークが既にインストール／セットアップされていて、すべての必須なサービスが正常に動作し、利用するコネクションタイプが正しく登録されていることを前提にします。これらの作業を自分で行う場合は、「開発者リソース」ページを参照してください。

本マニュアルは主にManifoldCFのユーザインタフェースの使い方について説明します。ManifoldCFの標準的な設定では、ブラウザで次のURLを開きます:`http://my-server-name:8345/mcf-crawler-ui`。もちろん、これはシステム毎に異なるでしょう。ご自身の環境での適切なURLについてはシステム管理者に問い合わせてください。

ManifoldCFのUIはFirefox及び色々なバージョンのIEでテストされています。他のWebブラウザを利用される場合は、正しく動作しない可能性もあります。ブラウザの非互換性の問題を見つけた場合は、システム管理者に連絡してください。

フレームワークのユーザインタフェースを初めて表示した時に、まずログインするように求められます:

WebページからURLを開くと、以下のようなページが表示されます:

左側に選択できるメニューが表示されます。右側には挨拶メッセージが表示されていますが、左側のメニューから項目を選択すると、右側に表示される内容が変わります。本格的に始める前に、下のメニューの説明に目を通してManifoldCFの機能の概要を理解することをお勧めします。

### 1.1 出力コネクションの定義

フレームワークのUIの左側のメニューには、出力コネクションの一覧へのリンクがあります。出力コネクションとは、様々なリポジトリから取得したドキュメントを書き込むことのできるシステムや場所へのコネクションです。これは多くの場合、検索エンジンです。

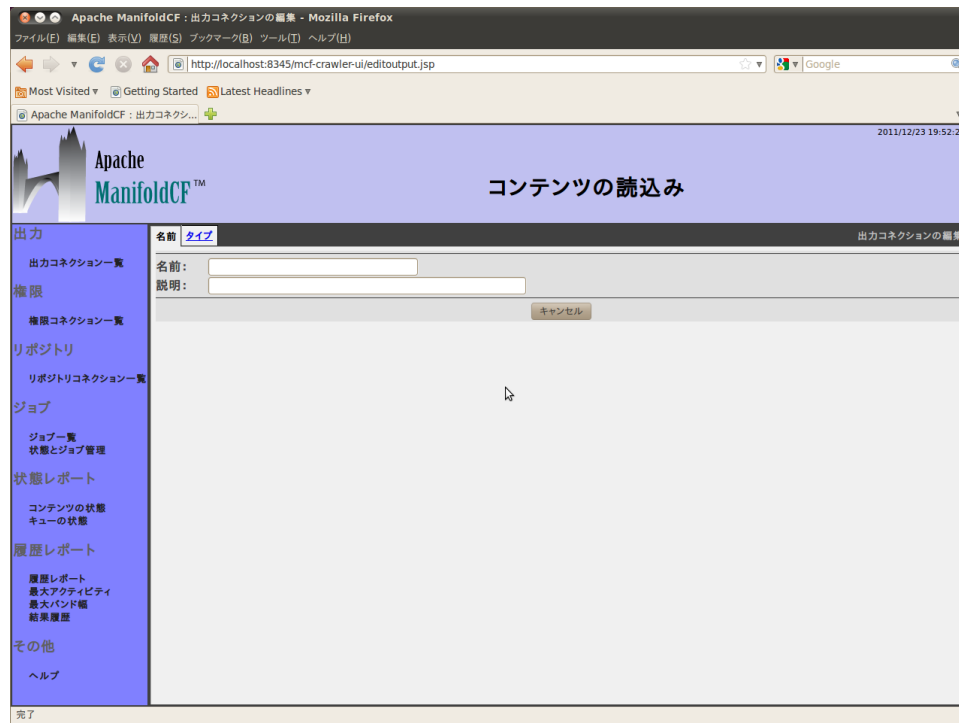
ジョブにはすべて出力コネクションを指定することが必要です。出力コネクションを作成するには、左側メニューから「出力コネクション一覧」を選択してください。以下のような画面が表示されます：

名前	説明	コネクションタイプ	最大値
Mega		MetaCarta GTS	10
NullOutput		Null	10
OpenServer		OpenSearchServer	10
output1		Solr	10
Solr	Solr Connection	Solr	10

新しい出力コネクションを追加

初めて利用する場合は、出力コネクションが表示されていないかもしれません。出力コネクションが定義されている場合は、この画面に一覧表示されます。出力コネクションを表示したり、編集したり、削除したりするためのリンクも合わせて表示されています。新しい出力コネクションを定義する場合は、一覧の下に「新しい出力コネクションを追加」リンクをクリックしてください。以下のような画面が表示されます：





上に表示されるタブはそれぞれ出力コネクションの様々な見方を表しています。各タブでコネクションの様々な属性を編集できます。選択したコネクションタイプに依存して、表示されるタブは異なります。

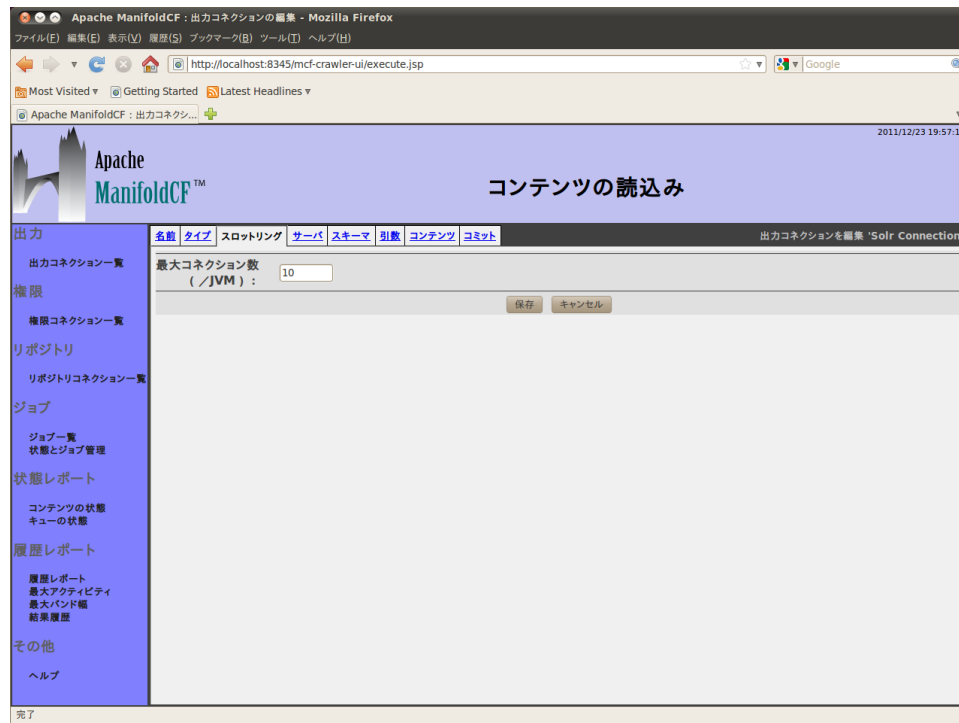
まず、コネクションに付ける名前と説明を入力してください。出力コネクション名は一意でなければなりません。また、一旦設定すると変更できませんので注意してください。名前は32文字以内、説明は255文字以内で設定してください。入力したら、「タイプ」タブを選択してください。コネクションのタイプタブが表示されます：



プルダウンボックスの中に表示される出力コネクションタイプの一覧とそれぞれの名前は、システムインテグレータにより決められます。以降の節で各出力コネクションタイプに対するタブの設定内容を説明します。

出力コネクションタイプを選択して、下の「次へ」ボタンを押してください。選択された出力コネクションに対応するタブが表示されます。また、下に「保存」ボタンも表示されます。コネクションを作成する場合は、この「保存」ボタンを必ず押してください。代わりに「キャンセル」ボタンを押せばコネクションは作成されません。(左に表示されるメニュー項目を選択しても同じことです。)

出力コネクションにはすべて「スロットリング」タブがあります。選択すると以下のようなページが表示されます：



このタブには項目は一つのみです: システムがこの出力コネクション用に利用できる最大のコネクション数です。この制限は、システムが過負荷になったり、またある場合にはライセンスの制限を超えたりするのを防ぐのに使えます。逆に、値を大きくすると、スループットが上がります。デフォルト値は10ですが、出力コネクションのタイプすべてに対して最適というわけではないかもしれません。より正確な値については、以降の各種の出力コネクションタイプを説明している節を参照してください。

各コネクションタイプに対応するタブの説明は、対象の出力コネクションタイプを説明している節を参照してください。

コネクションを保存すると、設定したコネクションの内容の要約画面が表示されます。(どのコネクションタイプを選んだかによって詳細は多少異なりますが)以下ようになります:



要約画面には接続の状態も表示されます。接続が正しく設定された場合は、状態は「Connection working」(正常)と表示されます。設定に間違いがある場合は、代わりに接続タイプに固有の診断メッセージが表示されます。そうなったら、インフラを修理するか接続の設定を適切に編集するかして、問題を修正してください。

下に「更新」、「編集」、「削除」、「関連するすべてのコンテンツの再読み込み」、「関連するすべてのコンテンツの削除」の5つのボタンがあります。各ボタンの目的を順に詳しく説明します。

「更新」ボタンは、単に出力接続の画面をリロードして、接続の状態を更新します。出力接続が接続されている外部のシステムに変更を加え、その変更が接続が成功するか否かに影響する場合に、このボタンを使います。

「編集」ボタンを押すと、前に戻って接続パラメータの編集ができます。接続の属性または仕様を変更したい場合にこのボタンを使います。

「削除」ボタンを押すと、接続が削除できます。利用可能な出力接続の一覧から接続を削除したい場合に、このボタンを使います。ManifoldCFでジョブから参照中の接続は削除できないことに注意してください。

「関連するすべてのコンテンツの再読み込み」ボタンを押すと、この接続で現在索引が作成されているコンテンツすべての記録されている版を無効にします。これは頻繁に使うボタンではありません。出力接続に書かれているシステムの構成を変更して、その結果、コンテンツすべての索引の再作成が必要になった時にクリックしてください。

「関連するすべてのコンテンツの削除」ボタンを押すと、この接続に関連して行なわれた索引に関する情報をすべてManifoldCFから削除します。これも頻繁に使うボタンではありません。対象のリポジトリから出力接続に書かれている索引をすべて削除した時に、クリックしてください。

## 1.2 変換コネクションの定義

フレームワークのUIの左側のメニューには変換コネクションの一覧へのリンクがあります。変換コネクションとは様々なリポジトリから取得したドキュメントを操作可能なエンジンへのコネクションです。通常メタデータ抽出やマッピングを伴います。

ジョブは必ずしも変換コネクションを指定する必要はありません。多くの場合、最終目的の検索エンジンには、付属のデータ変換パイプラインがあります。しかし、そのようなデータ抽出や変換が利用できない場合、ManifoldCFは内部的にそのような処理を引き受ける方法を提供します。

変換コネクションを作成するには、左側メニューから「変換コネクション一覧」を選択してください。以下のような画面が表示されます：

名前	説明	コネクションタイプ	最大値
1	Tika content extractor		10

初めて利用する場合は、変換コネクションは定義されていないかもしれませんが。変換コネクションが定義されている場合は、この画面に一覧表示されます。変換コネクションを表示したり、編集したり、削除したりするためのリンクも合わせて表示されています。新しい変換コネクションを定義する場合は、一覧の下に「新しい変換コネクションを追加」リンクを選択してください。選択すると以下のようなページが表示されます：

上に表示されるタブはそれぞれ変換コネクションの様々な見方を表しています。各タブでコネクションの様々な属性を編集できます。選択したコネクションタイプに依存して、表示されるタブは異なります。

まず、コネクションに付ける名前と説明を入力してください。変換コネクション名は一意でなければなりません。また、一旦設定すると変更できませんので注意してください。名前は32文字以内、説明は255文字以内で設定してください。入力したら、「タイプ」タブを選択してください。コネクションのタイプタブが表示されます：

プルダウンボックスの中に表示される変換コネクションタイプの一覧とそれぞれの名前は、システムインテグレータにより決められます。以降の節で各変換コネクションタイプに対するタブの設定内容を説明します。

変換コネクションタイプを選択して、下の「次へ」ボタンを押してください。選択された変換コネクションに対応するタブが表示されます。また、ページの下に「保存」ボタンも表示されます。コネクションを作成する場合はこの「保存」ボタンを必ず選択してください。代わりに「キャンセル」ボタンを押せばコネクションは作成されません。(左に表示されるメニュー項目を選択しても同じことです。)

変換コネクションタイプにはすべて「スロットリング」タブがあります。選択すると以下のようページが表示されます:

このタブには項目は一つのみです: システムがこの変換コネクション用に利用できる最大のコネクション数です。この制限は、システムが過負荷になったり、またある場合にはライセンスの制限を超えたりするのを防ぐのに使えます。逆に、値を大きくすると、スループットが上がります。デフォルト値は10ですが、変換コネクションのタイプすべてに対して最適というわけではないかもしれません。より正確な値については、以降の各種の変換コネクションタイプを説明している節を参照してください。

各コネクションタイプに対応するタブの説明は、対象の変換コネクションタイプを説明している節を参照してください。

コネクションを保存すると、設定したコネクションの内容の要約画面が表示されます。(どのコネクションタイプを選んだかによって詳細は多少異なりますが)以下ようになります:

要約画面にはコネクションの状態も表示されます。コネクションが正しく設定された場合は、状態は「Connection working」(正常)と表示されます。設定に間違いがある場合は、代わりにコネクションタイプに固有の診断メッセージが表示されます。そうなったら、インフラを修理するかコネクションの設定を適切に編集するかして、問題を修正してください。

下に「更新」、「編集」、「削除」の3つのボタンがあります。各ボタンの目的を順に詳しく説明します。

「更新」ボタンは、単に変換コネクションの画面をリロードして、コネクションの状態を更新します。変換コネクションが接続されている外部のシステムに変更を加え、その変更がコネクションが成功するか否かに影響する場合に、このボタンを使います。

「編集」ボタンを押すと、前に戻ってコネクションパラメータの編集ができます。コネクションの属性または仕様を変更したい場合にこのボタンを使ってください。



「削除」ボタンを押すと、コネクションが削除できます。利用可能な変換コネクションの一覧からコネクションを削除したい場合に、このボタンを使います。ManifoldCFでジョブから参照中のコネクションは削除できないことに注意してください。

### 1.3 権限グループの定義

フレームワークのUIの左側メニューには権限グループ一覧へのリンクがあります。権限グループとは、リポジトリの取得対象ドキュメントに対し協調してセキュリティを提供するための権限の集合です。例えば、クレームベース認証のSharePoint 2010リポジトリに含まれるドキュメントはSharePoint自身、Active Directory(アクティブディレクトリ)、その他により認証されます。従って、このようなSharePointリポジトリのドキュメントはSharePointネイティブ権限、SharePoint Active Directory権限、その他SharePointクレームベース権限をメンバーとする権限グループを参照する必要があります。ただしほとんどの場合、一つの権限のみで構成された権限グループでもリポジトリのセキュリティ保護は十分です。

権限コネクションを定義する際に権限グループを選択する必要があるため、権限コネクションを定義する前に権限グループを定義してください。権限グループが未定義の場合、権限コネクションを作成できません。しかし、権限コネクションを定義する際に間違った権限グループを選択した場合、後で修正することは可能です。

任意のリポジトリコネクションを作成する前に、権限グループを作成しておくこともお勧めします。その理由は、リポジトリコネクションもセキュアなドキュメントのために逆に権限グループを参照する場合があるからです。後でリポジトリコネクションと権限グループの関係を変更することも可能ですが、関連ジョブが次に実行される時に多くのドキュメントが再インデックス付けされる場合があります。

権限グループを作成するには、左側メニューから「権限グループ一覧」を選択してください。次のような画面が表示されます：

権限グループが定義されている場合は、一覧表示されます。権限グループの左に表示される「表示」、「編集」、「削除」リンクを選択して設定内容を表示、編集、削除することができます。新しい権限グループを定義する場合は、一覧の下の「新しい権限グループを追加」リンクを選択してください。選択すると以下のようなページが表示されます：

上に表示されるタブはそれぞれ権限グループの様々な見方を表しています。権限グループには「名前」タブしか表示されません。

権限グループに名前と説明を入力してください。権限グループ名は一意である必要があります。また、一旦設定すると変更できませんので注意してください。名前は32文字以内、説明は255文字以内で設定してください。入力した後に「保存」ボタンをクリックしてください。権限グループを作成または更新した際は必ず「保存」ボタンをクリックしてください。代わりに「キャンセル」ボタンを押せば新しい権限グループは作成されません。(左に表示されるメニュー項目を選択しても同じことです。)

権限グループを保存した後、要約画面にグループ情報が表示されます。その後、本権限グループに属する任意の権限コネクション、または本権限グループを参照するリポジトリコネクションを作成してください。

#### 1.4 リポジトリコネクションの定義

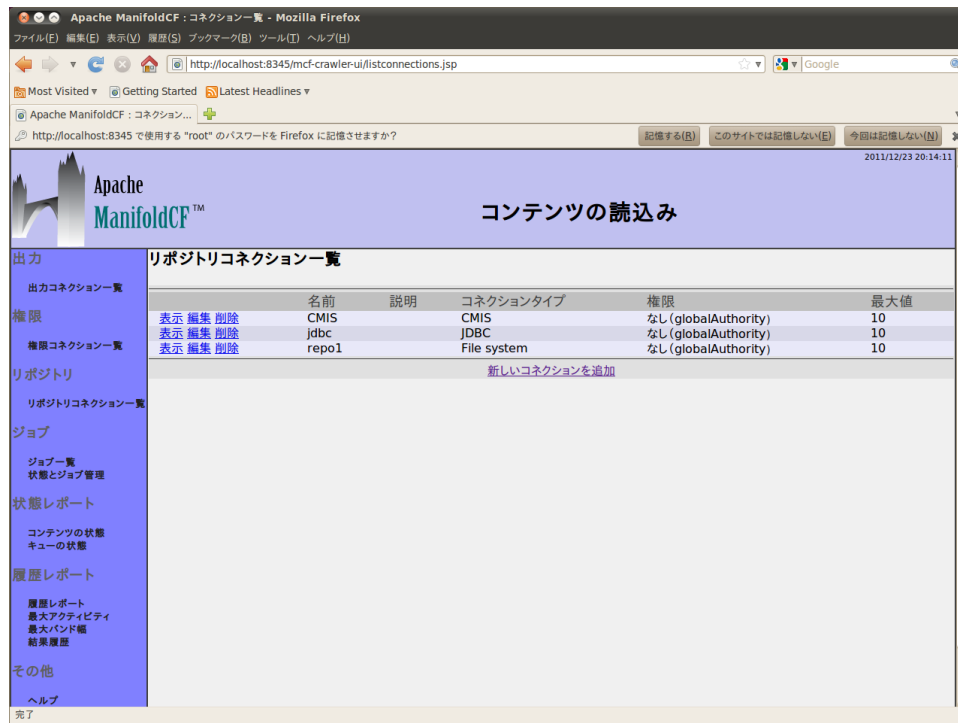
フレームワークのUIの左側メニューには、リポジトリコネクションの一覧へのリンクがあります。リポジトリコネクションとは、索引を作成するコンテンツを保管しているリポジトリシステムへのコネクションです。

そこからコンテンツを読み込むので、ジョブはすべてリポジトリコネクションを指定する必要があります。そのため、コンテンツから索引を作成する前に、リポジトリコネクションを作成することが必要です。

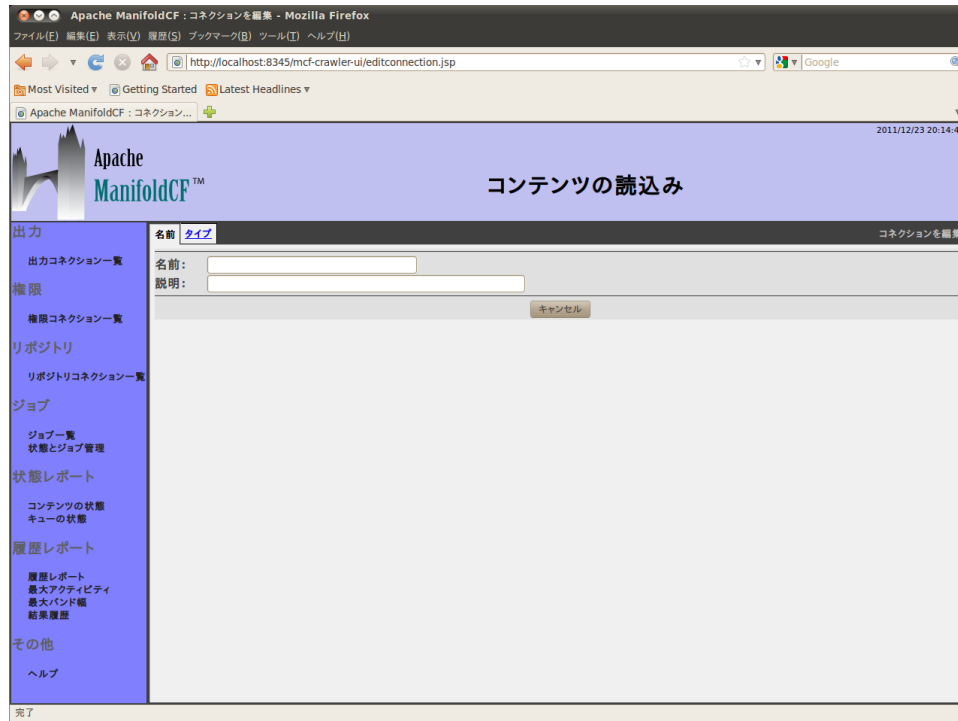
リポジトリコネクションに関連する権限グループを指定することもできます。ここで指定された権限グループは、リポジトリコネクションで読み込むコンテンツのセキュリティ環境を決定します。クローラーを実行した後にもリポジトリコネクションに対応した権限グループを変更することもできますが、この場合は誰でも検索可能であるためリポジトリコネクションが対象とするすべてのコンテンツの索引を再構成することが必要です。そのため、リポジトリコネクションを定義する前に権限グループを設定することを推奨します。

リポジトリコネクションを作成するには左側のメニューから「リポジトリコネクション一覧」を選択します。次のような画面が表示されます：





初めて利用する場合は、リポジトリコネクションは表示されていないかもしれません。リポジトリコネクションが定義されている場合は、この画面に一覧表示されます。リポジトリコネクションを表示したり、編集したり、削除したりするためのリンクも合わせて表示されています。新しいリポジトリコネクションを定義する場合は、一覧の下の「新しいコネクションを追加」リンクをクリックしてください。以下のような画面が表示されます：



上に表示されるタブはそれぞれリポジトリコネクションの様々な見方を表しています。各タブでコネクションの様々な属性を編集できます。選択したコネクションタイプに依存して、表示されるタブは異なります。

まず、コネクションに付ける名前と説明を入力してください。リポジトリコネクション名は一意でなければなりません。また、一旦設定すると変更できませんので注意してください。名前は32文字以内、説明は255文字以内で設定してください。入力したら、「タイプ」タブを選択してください。コネクションのタイプタブが表示されます：

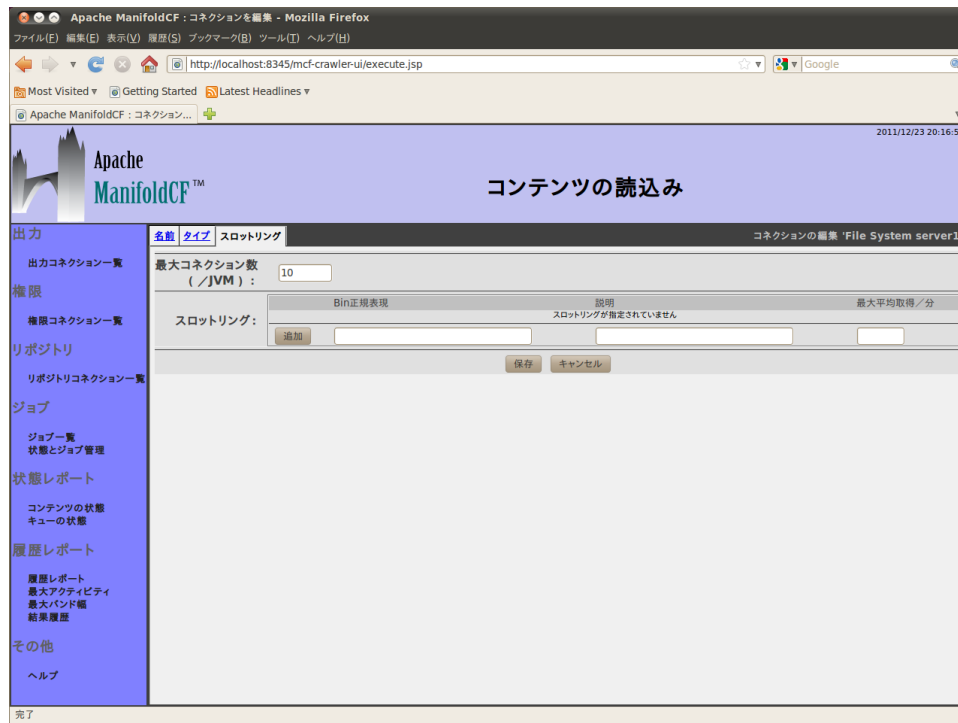


プルダウンボックスの中に表示されるリポジトリコネクションタイプの一覧とそれぞれの名前は、システムインテグレータにより決められます。以降の節で各リポジトリコネクションタイプに対するタブの設定内容を説明します。

この時点でリポジトリから読み込むコンテンツを保護するのに使う権限グループを選択することもできます。この段階より前に権限グループの権限コネクションを定義する必要はありませんが、そうしないと索引を作成したコンテンツを検索できません。

リポジトリコネクションタイプと(ご希望なら)権限グループを選択した後は下の「次へ」ボタンをクリックしてください。選択されたりポジトリコネクションに対応するタブが表示されます。また、ページの下に「保存」ボタンも表示されます。コネクションを作成する場合はこの「保存」ボタンを必ず押してください。代わりに「キャンセル」ボタンを押せばコネクションは作成されません。(左に表示されるメニュー項目を選択しても同じことです。)

リポジトリコネクションにはすべて「スロットリング」タブがあります。選択すると以下のようなページが表示されます：



このタブには二つの項目があります。まず一つ目は、システムがこのリポジトリコネクションに利用できる最大のコネクション数です。この制限は、システムが過負荷になったり、またある場合にはライセンスの制限を超えたりするのを防ぐのに使えます。逆に、値を大きくすると、スループットが上がります。デフォルト値は10ですが、リポジトリコネクションのタイプすべてに対して最適というわけではないかもしれません。より正確な値については、以降の各種のリポジトリコネクションタイプを説明している節を参照してください。二つ目は、クローラーがこのコネクションを利用して平均的にどれだけ早くコンテンツを読み込むかです。

コネクション毎に「スロットルbin」を設定することができます。スロットルbinとは、読み込み頻度を制限するリソースの名前です。(恐らく)個々のサーバは独立して頻度を制御することが必要でしょうから、Webコネクションはコンテンツのサーバ名をスロットルbinとして使います。

リポジトリコネクションの「スロットリング」タブではスロットリング設定を数に制限なく定義できます。各スロットリング設定は、スロットルbin集を表す正規表現、説明文、正規表現に一致するスロットルbin毎の1分あたりの平均読み込み頻度から構成されています。スロットルbinが1つ以上のスロットリング設定と一致した場合は、一番資源を利用しない読み込み設定が選ばれます。

一番簡単な正規表現は空の正規表現です。この場合はコネクションのスロットルbinすべてと一致します。コネクションにデフォルトのスロットル設定を指定する場合には、この方法を

使って設定することができます。平均読込み率を指定して、「追加」ボタンをクリックしてください。以下のようなスロットリングタブが表示されます：



スロットル設定を行わない場合は、コンテンツの読込みはスロットルされません。

各コネクションタイプに対応するタブの説明は、対象のリポジトリコネクションタイプを説明している節を参照してください。

コネクションを保存すると、設定したコネクションの内容の要約画面が表示されます。(どのコネクションタイプを選んだかによって詳細は多少異なりますが)以下ようになります：

Outputs	View Repository Connection Status		
List Output Connections			
Authorities	Name:	files	Description:
List Authority Groups	Connection type:	File system	Max connections: 10
List User Mapping Connections	Authority group:	None (global authority)	
List Authority Connections	Throttling:	Bin regular expression	Description
Repositories		No throttles	Max avg fetches/min
List Repository Connections	Connection status: Connection working		
Jobs	Refresh Edit Delete Clear all related history		
List all Jobs			

要約画面にはコネクションの状態も表示されます。コネクションが正しく設定された場合は、状態は「Connection working」(正常)と表示されます。設定に間違いがある場合は、代わりにコネクションタイプに固有の診断メッセージが表示されます。そうなったら、インフラを修理するかコネクションの設定を適切に編集するかして、問題を修正してください。

下に「更新」、「編集」、「削除」、「すべての関連履歴をクリア」の4つのボタンがあります。各ボタンの目的を順に詳しく説明します。

「更新」ボタンは、単にリポジトリコネクションの画面をリロードして、コネクションの状態を更新します。リポジトリコネクションが接続されている外部のシステムに変更を加え、その変更がコネクションが成功するか否かに影響する場合に、このボタンを使います。

「編集」ボタンを押すと、前に戻ってコネクションパラメータの編集ができます。コネクションの属性または仕様を変更したい場合にこのボタンを使います。

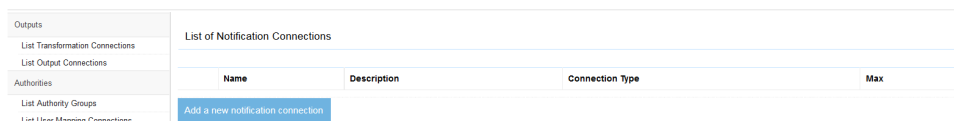
「削除」ボタンを押すと、コネクションが削除できます。利用可能なリポジトリコネクションの一覧からコネクションを削除したい場合に、このボタンを使います。ManifoldCFでジョブから参照中のコネクションは削除できないことに注意してください。

「すべての関連履歴をクリア」ボタンを押すと、現在のリポジトリコネクションに関連する履歴データをすべて削除します。これは頻繁に使うボタンではありません。履歴データは、「履歴レポート」などの、レポートを作るのに使われます。現時刻から指定された期間よりも古い履歴を自動的に削除するようにManifoldCFを設定する方法があります。しかし、今すぐ履歴をすべて削除したい場合には、このボタンを使います。

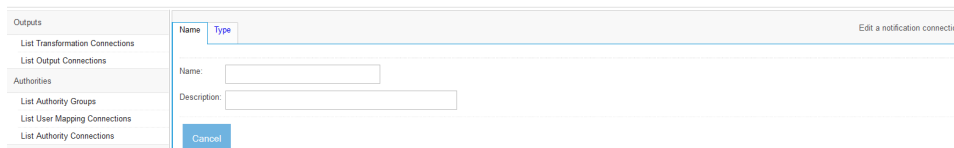
## 1.5 通知コネクションの定義

フレームワークのUIの左側メニューには、通知コネクションの一覧へのリンクがあります。通知コネクションとは、メールやテキストメッセージのような通知メッセージを生成するエンジンへのコネクションです。通知メッセージは、特にジョブの完了や意図せぬ終了を通知します。

ジョブには1つ、もしくはそれ以上の通知コネクションを指定できます。通知コネクションを作成するには、左側メニューから「通知コネクション一覧」を選択してください。以下のような画面が表示されます：



初めて利用する場合は、通知コネクションが表示されていないかもしれません。通知コネクションが定義されている場合は、この画面に一覧表示されます。通知コネクションを表示したり、編集したり、削除したりするためのリンクも合わせて表示されています。新しい通知コネクションを定義する場合は、一覧の下に「新しい通知コネクションを追加」リンクをクリックしてください。以下のような画面が表示されます：



上に表示されるタブはそれぞれ通知コネクションの様々な見方を表しています。各タブでコネクションの様々な属性を編集できます。選択したコネクションタイプに依存して、表示されるタブは異なります。

まず、コネクションに付ける名前と説明を入力してください。出力コネクション名は一意でなければなりません。また、一旦設定すると変更できませんので注意してください。名前は32文字以内、説明は255文字以内で設定してください。入力したら、「タイプ」タブを選択してください。コネクションのタイプタブが表示されます：

プルダウンボックスの中に表示される通知コネクションタイプの一覧とそれぞれの名前は、システムインテグレータにより決められます。以降の節で各出力コネクションタイプに対するタブの設定内容を説明します。

通知コネクションタイプを選択して、下の「次へ」ボタンを押してください。選択された通知コネクションに対応するタブが表示されます。また、下に「保存」ボタンも表示されます。コネクションを作成する場合は、この「保存」ボタンを必ず押してください。代わりに「キャンセル」ボタンを押せばコネクションは作成されません。(左に表示されるメニュー項目を選択しても同じことです。)

通知コネクションにはすべて「スロットリング」タブがあります。選択すると以下のようなページが表示されます：

このタブには指定できる項目は一つのみです：システムがこの通知コネクション用に利用できる最大のコネクション数です。この制限は、システムが過負荷になったり、またある場合にはライセンスの制限を超えたりするのを防ぐのに使えます。逆に、値を大きくすると、スループットが上がります。デフォルト値は10ですが、通知コネクションのタイプすべてに対して最適というわけではないかもしれません。より正確な値については、以降の各種の通知コネクションタイプを説明している節を参照してください。

各コネクションタイプに対応するタブの説明は、対象の通知コネクションタイプを説明している節を参照してください。

コネクションを保存すると、設定したコネクションの内容の要約画面が表示されます。(どのコネクションタイプを選んだかによって詳細は多少異なりますが)以下のようになります：



Outputs	View Notification Connection Status	
List Transformation Connections		
List Output Connections		
Authorities		
List Authority Groups		
List User Mapping Connections		
List Authority Connections		
Repositories		
List Repository Connections		
List Notification Connections		
Jobs		

Name:	Notification	
Connection type:	E-Mail	
Protocol:	IMAP	
Host name:	localhost	
Port:		
User name:	admin	
Password:	*****	
Configuration properties:		
Connection status: Connection failed: Error connecting: Connection refused: connect		
<input type="button" value="Refresh"/> <input type="button" value="Edit"/> <input type="button" value="Delete"/>		

要約画面にはコネクションの状態も表示されます。コネクションが正しく設定された場合は、状態は「Connection working」(正常)と表示されます。設定に間違いがある場合は、代わりにコネクションタイプに固有の診断メッセージが表示されます。そうなったら、インフラを修理するかコネクションの設定を適切に編集するかして、問題を修正してください。

下に「更新」、「編集」、「削除」の3つのボタンがあります。各ボタンの目的を順に詳しく説明します。

「更新」ボタンは、単に通知コネクションの画面をリロードして、コネクションの状態を更新します。通知コネクションが接続されている外部のシステムに変更を加え、その変更がコネクションが成功するか否かに影響する場合に、このボタンを使います。

「編集」ボタンを押すと、前に戻ってコネクションパラメータの編集ができます。コネクションの属性または仕様を変更したい場合にこのボタンを使います。

「削除」ボタンを押すと、コネクションが削除できます。利用可能な通知コネクションの一覧からコネクションを削除したい場合に、このボタンを使います。ManifoldCFでジョブから参照中のコネクションは削除できないことに注意してください。

## 1.6 ユーザーマッピングコネクションの定義

フレームワークのUIの左側メニューにはユーザーマッピングコネクション一覧へのリンクがあります。ユーザーマッピングコネクションとは、ユーザ名を別ユーザ名にマッピングする方法を理解しているシステムへのコネクションです。例えば、LiveLinkを用いた文書安全対策を実施したいが、Active Directoryユーザー名しか持っていない場合、LiveLink権限で使うためのアクセストークンを発見する前に、Active Directoryユーザー名を該当するLiveLinkユーザ名にマッピングする必要があります。

実用的なユーザーマッピングコネクションはすべて他のシステムにアクセスすることが必要というわけではありません。ManifoldCFには、例えば、ユーザ名文字列を独自に正規表現で処理するための正規表現ユーザーマッパーが付いています。また、ユーザマッピングはほとんどではないにしても多くの権限において必要ありません。作成予定の権限が同じユーザ名で完全に操作できて、かつそのユーザ名が検索時にManifoldCFの権限サブレットにおいて利用可能である場合、ユーザーマッピングコネクションは必要ありません。

権限コネクションを設定する前にマッピングコネクションを定義すべきです。理由は、権限コネクションで予め定義されたマッピングコネクションを指定する場合があるからです。同様な



理由から、ユーザ名を処理したい場合もマッピングコネクションを定義しておくのが便利です。初回で正しく設定できなくても、後で戻って修正可能です。

ユーザーマッピングコネクションを作成するには、左側メニューから「ユーザーマッピングコネクション一覧」を選択してください。次のような画面が表示されます：

初めて利用する場合は、ユーザーマッピングコネクションは定義されていないかもしれませんが、ユーザーマッピングコネクションが定義されている場合は、この画面に一覧表示されます。ユーザーマッピングコネクションを表示したり、編集したり、削除したりするためのリンクも合わせて表示されています。新しいユーザーマッピングコネクションを定義する場合は、一覧の下の「新しいコネクションを追加」リンクを選択してください。選択すると以下のようなページが表示されます：

上に表示されるタブはそれぞれマッピングコネクションの様々な見方を表しています。各タブでコネクションの様々な属性を編集できます。選択したコネクションタイプに依存して、表示されるタブは異なります。

まず、コネクションに付ける名前と説明を入力してください。ユーザーマッピングコネクション名は一意でなければなりません。また、一旦設定すると変更できませんので注意してください。名前は32文字以内、説明は255文字以内で設定してください。入力したら、「タイプ」タブを選択してください。コネクションのタイプタブが表示されます：

プルダウンボックスの中に表示されるマッピングコネクションタイプの一覧とそれぞれの名前は、システムインテグレータにより決められます。以降の節でManifoldCFに含まれる各マッピングコネクションタイプに対するタブの設定内容を説明します。

マッピングコネクションタイプを選択して、下の「次へ」ボタンを選択してください。選択されたユーザーマッピングコネクションに対応するタブが表示されます。また、ページの下に「保存」ボタンも表示されます。コネクションを作成する場合はこの「保存」ボタンを必ず選択してください。代わりに「キャンセル」ボタンを押せばコネクションは作成されません。(左に表示されるメニュー項目を選択しても同じことです。)

マッピングコネクションにはすべて「条件」タブがあります。このタブで、(もしあれば)このコネクションに先立って実行すべきマッピングコネクションを指定できます。タブは以下の通りです:

出力	名前	タイプ	条件	スロットリング	User Mapping	マッピングを編集
変換コネクション一覧	前提ユーザーマッピング: (条件無し)					
出力コネクション一覧						
権限	保存 キャンセル					
権限グループ一覧						
ユーザーマッピングコネクション一覧						

注意: ループ関係にある複数条件を指定することは絶対避けてください。そのため、ManifoldCFはプルダウンリストにループに陥るようなユーザーマッピングコネクションを表示しないようにしています。

マッピングコネクションタイプにはすべて「スロットリング」タブがあります。選択すると以下のようなページが表示されます:

出力	名前	タイプ	条件	スロットリング	User Mapping	マッピングを編集
変換コネクション一覧	最大コネクション数: 10					
出力コネクション一覧						
権限	保存 キャンセル					
権限グループ一覧						
ユーザーマッピングコネクション一覧						

このタブには項目は一つのみです: システムがこのユーザーマッピングコネクション用に利用できる最大のコネクション数です。この制限は、システムが過負荷になったり、またある場合にはライセンスの制限を超えたりするのを防ぐのに使えます。逆に、値を大きくすると、スループットが上がります。デフォルト値は10ですが、マッピングコネクションのタイプすべてに対して最適というわけではないかもしれません。より正確な値については、以降の各種のユーザーマッピングコネクションタイプを説明している節を参照してください。

各コネクションタイプに対応するタブの説明は、対象のマッピングコネクションタイプを説明している節を参照してください。

コネクションを保存すると、設定したコネクションの内容の要約画面が表示されます。(どのコネクションタイプを選んだかによって詳細は多少異なりますが)以下ようになります:

出力	マッピングコネクション状態の表示	
変換コネクション一覧	名前:	test
出力コネクション一覧	マッピングタイプ:	Regular expression mapper
権限	前提ユーザーマッピング:	条件無し
権限グループ一覧	User name regular expression:	^(?!@)([A-Za-z0-9_!@#\$%^&*]{1,64})\$
ユーザーマッピングコネクション一覧	User expression:	\$2/\$1
権限コネクション一覧	コネクション状態:	Connection working
リポジトリ	更新 編集 削除	
リポジトリコネクション一覧		

要約画面にはコネクションの状態も表示されます。コネクションが正しく設定された場合は、状態は「Connection working」(正常)と表示されます。設定に間違いがある場合は、代わりにコネクションタイプに固有の診断メッセージが表示されます。そうなったら、インフラを修理するかコネクションの設定を適切に編集するかして、問題を修正してください。

下に「更新」、「編集」、「削除」の3つのボタンがあります。各ボタンの目的を順に詳しく説明します。

「更新」ボタンは、単にマッピングコネクションの画面をリロードして、コネクションの状態を更新します。マッピングコネクションが接続されている外部のシステムに変更を加え、その変更がコネクションが成功するか否かに影響する場合に、このボタンを使います。

「編集」ボタンを押すと、前に戻ってコネクションパラメータの編集ができます。コネクションの属性または仕様を変更したい場合にこのボタンを使います。

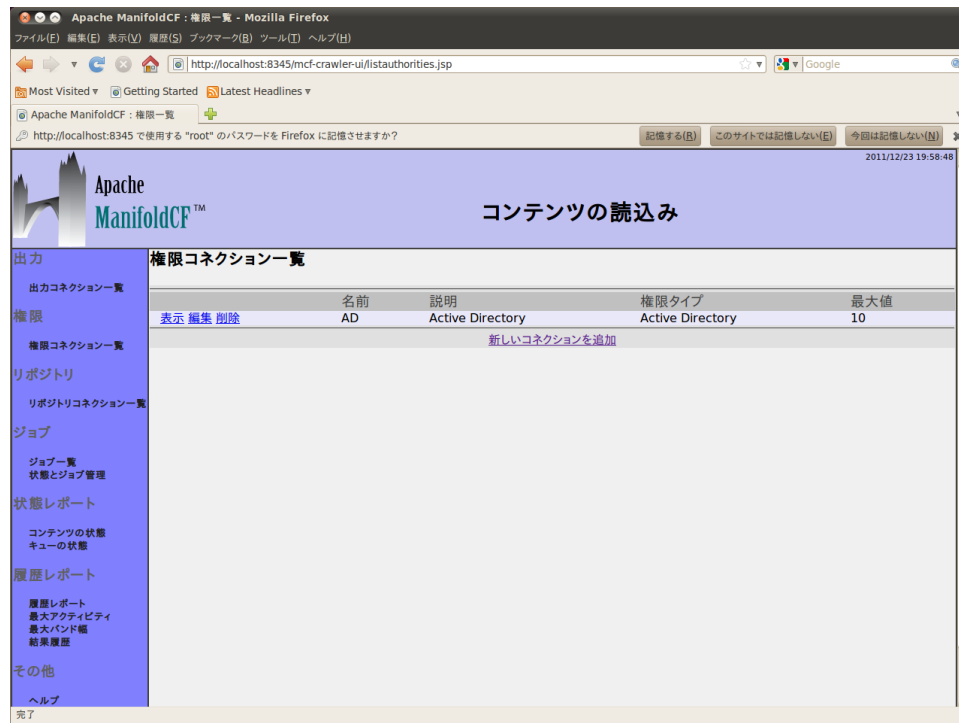
「削除」ボタンを押すと、コネクションが削除できます。利用可能なマッピングコネクションの一覧からコネクションを削除したい場合に、このボタンを使います。ManifoldCFでジョブから参照中のコネクションは削除できないことに注意してください。

### 1.7 権限コネクションの定義

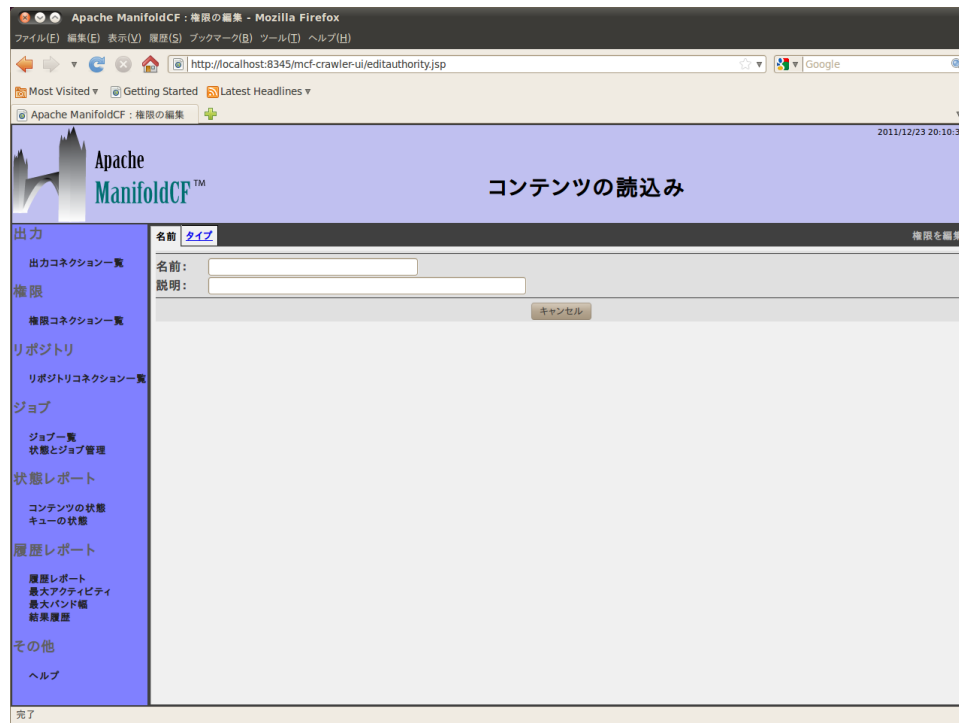
フレームワークのUIの左側メニューに権限コネクション一覧へのリンクがあります。権限コネクションとは、特定のセキュリティ環境を定義するシステムへのコネクションです。例えば、Active Directoryで保護されているファイルの索引を作る場合は、Active Directory権限コネクションを定義します。

特定の権限コネクションのみが特定のリポジトリコネクションタイプと互換性があることに注意してください。どのように使われることを想定しているのかを理解するには、この文書の対応するリポジトリタイプの詳細を読んでください。索引を作りたいコンテンツが誰からも見えることを気にしないのであれば、権限は必要ありません。例えば、ウェブ、RSS、Wikiをクロールする場合が当てはまるでしょう。しかしながら、それ以外リポジトリには独自のセキュリティ機構がある場合がほとんどです。

権限コネクションを作成するには、左側メニューから「権限コネクション一覧」を選択してください。以下のような画面が表示されます：

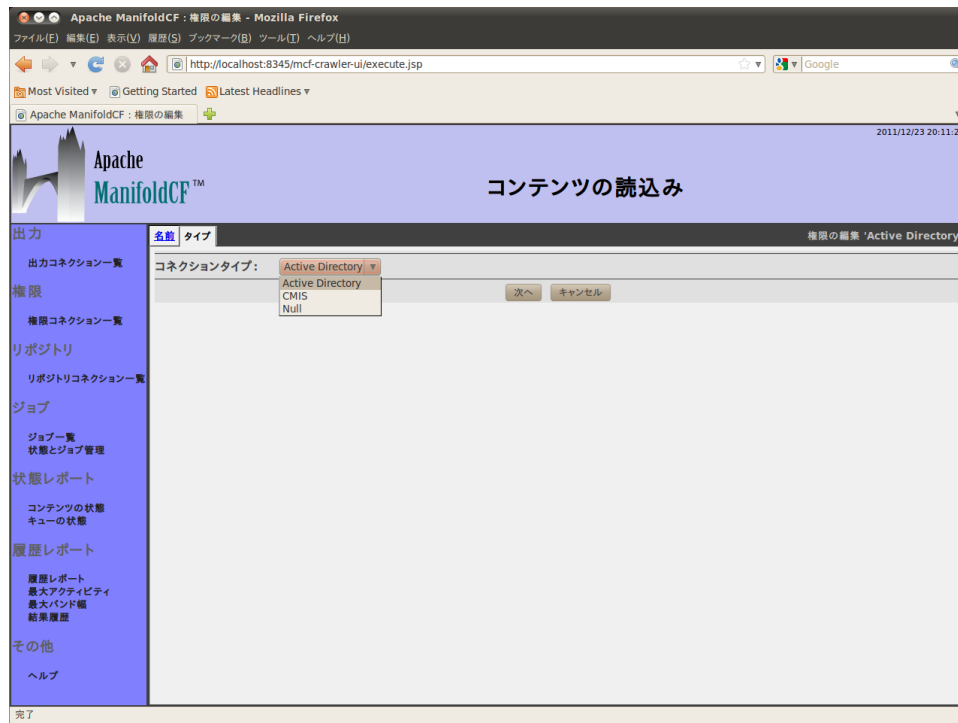


初めて利用する場合は、権限コネクションが表示されていないかもしれません。権限コネクションが定義されている場合は、この画面に一覧表示されます。出力コネクションを表示したり、編集したり、削除したりするためのリンクも合わせて表示されています。新しい権限コネクションを定義する場合は、一覧の下に「新しいコネクションを追加」リンクをクリックしてください。以下のような画面が表示されます：



上に表示されるタブはそれぞれ権限コネクションの様々な見方を表しています。各タブでコネクションの様々な属性を編集できます。選択したコネクションタイプに依存して、表示されるタブは異なります。

まず、コネクションに付ける名前と説明を入力してください。権限コネクション名は一意でなければなりません。また、一旦設定すると変更できませんので注意してください。名前は32文字以内、説明は255文字以内で設定してください。入力したら、「タイプ」タブを選択してください。コネクションのタイプタブが表示されます：



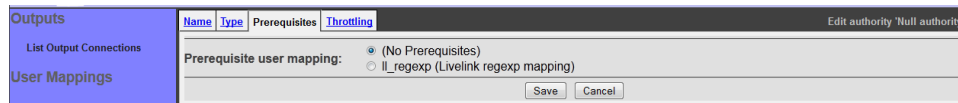
プルダウンボックスの中に表示される権限コネクションタイプの一覧とそれぞれの名前は、システムインテグレータにより決められます。以降の節で各権限コネクションタイプに対するタブの設定内容を説明します。

このタブでは作成している権限コネクションが属する権限グループも選ばなければなりません。プルダウンメニューから適切な権限グループを選んでください。

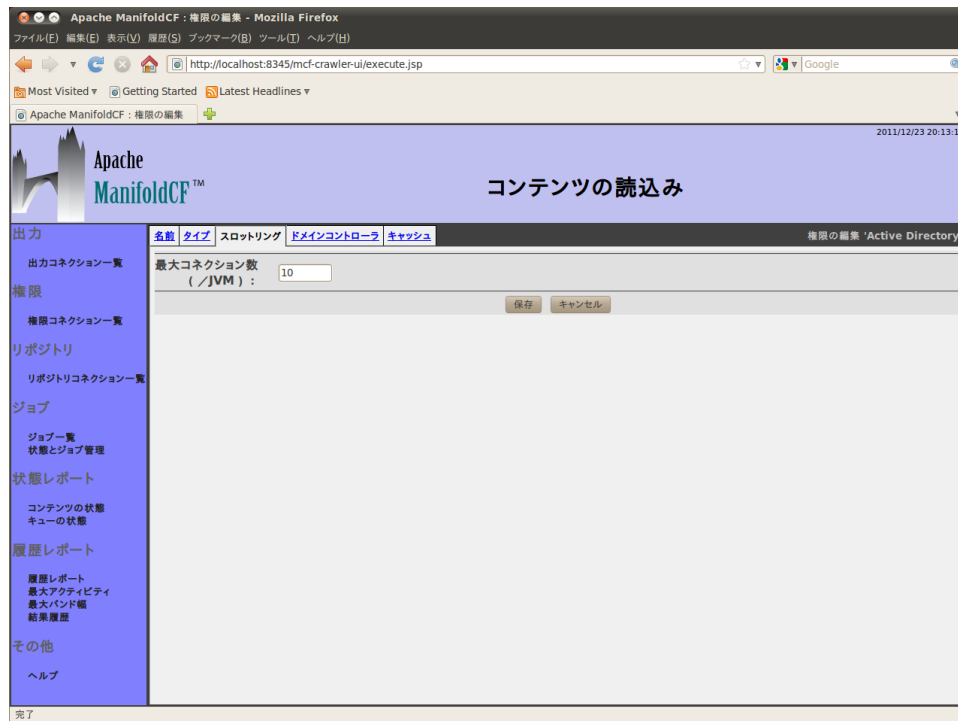
デフォルトでない認証ドメインを選ぶこともできます。認証ドメインは権限コネクションにどのユーザーIDが関連するかを記述するものです。例えば、1人のユーザーがActive DirectoryのID、LiveLinkのIDおよびFaceBookのIDを持っているかもしれません。ある権限コネクションはそのIDの1つに対してだけ適切でしょう。利用可能な認証ドメインの一覧はシステムインテグレータにより決められます。

権限コネクションタイプ、権限グループ、必要なら認証ドメインを選択して、下の「次へ」ボタンを押してください。選択された権限コネクションに対応するタブが表示されます。また、下に「保存」ボタンも表示されます。コネクションを作成する場合は、この「保存」ボタンを必ず押してください。代わりに「キャンセル」ボタンを押せばコネクションは作成されません。(左に表示されるメニュー項目を選択しても同じことです。)

権限コネクションにはすべて「条件」タブがあります。このタブで、(もしあれば)このコネクションに先立って実行すべきマッピングコネクションを指定できます。タブは以下の通りです:



権限コネクションにはすべて「スロットリング」タブもあります。選択すると以下のようなページが表示されます:



このタブには項目は一つのみです: システムがこの権限コネクション用に利用できる最大のコネクション数です。この制限は、システムが過負荷になったり、またある場合にはライセンスの制限を超えたりするのを防ぐのに使えます。逆に、値を大きくすると、スループットが上がります。デフォルト値は10ですが、権限コネクションのタイプすべてに対して最適というわけではないかもしれません。より正確な値については、以降の各種の権限コネクションタイプを説明している節を参照してください。

各コネクションタイプに対応するタブの説明は、対象の権限コネクションタイプを説明している節を参照してください。

コネクションを保存すると、設定したコネクションの内容の要約画面が表示されます。(どのコネクションタイプを選んだかによって詳細は多少異なりますが)以下のようになります:



Outputs	View Authority Connection Status	
List Output Connections		
Authorities		
List Authority Groups	Name: MyAuthorityConnection	Description: My authority connection
List User Mapping Connections	Authority type: Null	Max connections: 10
List Authority Connections	Authority group: MyAuthorityGroup	Authorization domain:
Repositories	Prerequisite user mapping: No prerequisites	
List Repository Connections	Connection status: Connection working	
Jobs	<input type="button" value="Refresh"/> <input type="button" value="Edit"/> <input type="button" value="Delete"/>	
List all Jobs		

要約画面にはコネクションの状態も表示されます。コネクションが正しく設定された場合は、状態は「Connection working」(正常)と表示されます。設定に間違いがある場合は、代わりにコネクションタイプに固有の診断メッセージが表示されます。そうなったら、インフラを修理するかコネクションの設定を適切に編集するかして、問題を修正してください。

下に「更新」、「編集」、「削除」の3つのボタンがあります。各ボタンの目的を順に詳しく説明します。

「更新」ボタンは、単に権限コネクションの画面をリロードして、コネクションの状態を更新します。権限コネクションが接続されている外部のシステムに変更を加え、その変更がコネクションが成功するか否かに影響する場合に、このボタンを使います。

「編集」ボタンを押すと、前に戻ってコネクションパラメータの編集ができます。コネクションの属性または仕様を変更したい場合にこのボタンを使います。

「削除」ボタンを押すと、コネクションが削除できます。利用可能な権限コネクションの一覧からコネクションを削除したい場合に、このボタンを使います。

## 1.8 ジョブの作成

ManifoldCFの「ジョブ」とは、コンテンツの集まりを記述したものです。フレームワークのジョブは指定されたリポジトリコネクションからコンテンツを読み込み、0個以上の変換コネクションを使って変換し、指定された出力コネクションに書込みます。ジョブの内容とコンテンツの索引作成方法は、関連するリポジトリコネクションに依存します。コンテンツの索引作成方法は、関連する出力コネクションや指定された変換コネクションにも依存します。

ジョブはすべて1回以上、実行されます。ジョブが実行される度に、新規のコンテンツ及び変更されたコンテンツを出力コネクションに送る他にも、対象外になったコンテンツに付いても出力コネクションに通知します。コンテンツは2つの方法で対象外になります: コンテンツがリポジトリから削除された場合、コンテンツがもはや許可されたコンテンツの集まりに含まれていない場合。フレームワークはこの両方の場合に対応しています。

ジョブを削除すると、そのジョブに関連したコンテンツすべてが削除されたことを出力コネクションに通知します。ジョブはそのジョブに関連したコンテンツを表していますので、この動作は道理にかなっています。ジョブが削除された場合に、関連したコンテンツも削除されないと、そのコンテンツに親がなくなってしまう。(ManifoldCFのジョブはタスクにすぎないと決めてかかる人もいますが、間違った想定です。)



フレームワークでは複数のジョブで1つのコンテンツを読み込むことができます。1つ以上のジョブに関連したコンテンツは以下のように特別に処理されます:

- ジョブを削除すると、他ジョブの対象に含まれていないコンテンツの削除の通知が出力コネクションに送られます。
- 出力コネクションに通知が送られるコンテンツのバージョンはどのジョブが最後に実行されるかによります。

コンテンツが複数ジョブの対象の場合の処理は複雑なため、できる限りこのような状況は避けたほうがよいです。

ジョブの非継続的な実行は典型的には以下のようなステージで行なわれます:

1. ジョブの新規、変更、削除の開始点をキューに登録(「シーディング」)
2. コンテンツを読み込み、新コンテンツを発見、削除を検出
3. キューから読み込み対象外になったコンテンツを削除

ジョブを「継続的」に走らせることもできます。継続的に走っているジョブは中断されるまで実行を継続します。継続ジョブは以下のようなステージで実行されます:

1. ジョブの新規、変更、削除の開始点をキューに登録(「シーディング」)
2. 定期的に再シーディングしている間、コンテンツを読み込み、新コンテンツを発見、削除を検出。

継続ジョブは除外コンテンツをキューから削除することはできないことに注意してください。リポジトリから削除されたコンテンツをキューから外すことができます。

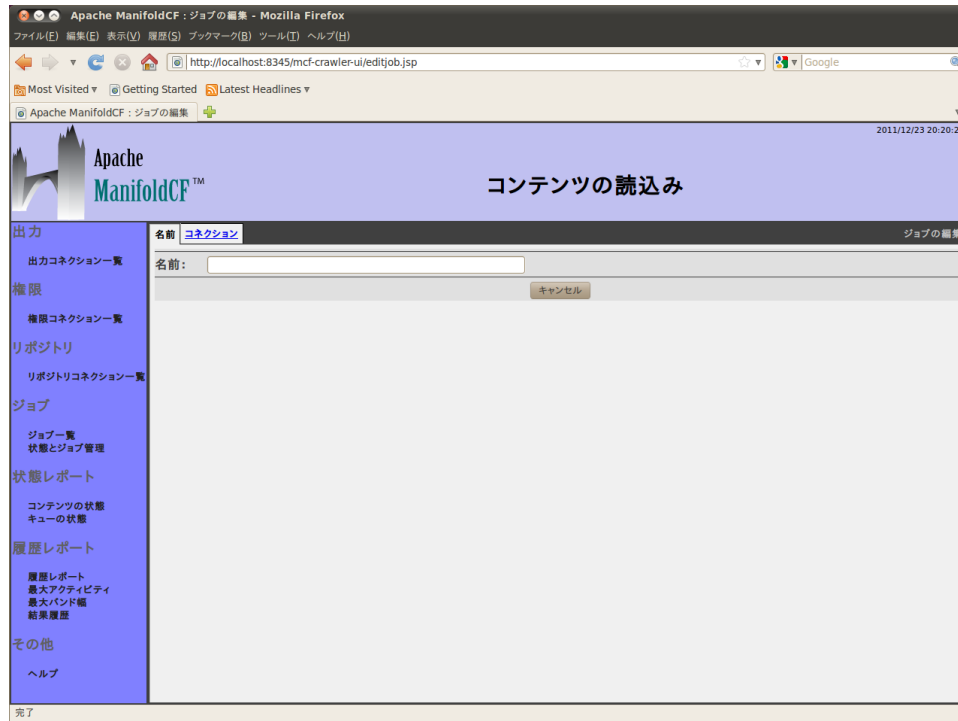
ジョブはユーザーが明示的に開始したときに即時に実行することも、ユーザーが指定したスケジュールで実行することも独立に設定できます。スケジュールで実行するよう設定した場合は、指定日時に開始することも、他ジョブが完了した後に実行するように設定することもできます。

ManifoldCFで並列実行可能なジョブ数に制限は設けられていません。

ジョブを作成する場合は、まず左メニューの「ジョブ一覧」リンクを選択します。次のような画面が表示されます:



定義されたジョブを表示、編集、削除するには、対応するリンクを選択してください。定義したジョブを複製することもできます。新規にジョブを定義する場合は、下の「新しいジョブの追加」リンクを選択してください。次のページが表示されます：



ジョブ名を入力してください。ジョブ名は一意である必要はありません。ただし、一意にした方が分かりやすいので一意にする事を推奨します。入力した後に「コネクション」タブを選択してください:



続いて、リポジトリコネクション名を選択してください。ジョブ定義を保存すると、選択したコネクションを変更することができなくなりますので注意してください。

プルダウンの出力を選び、条件パイプラインステージを選び、「出力を追加ボタン」を押して、出力コネクションを1つもしくは複数、追加してください。ジョブ定義を保存すると、出力を削除することができなくなりますので注意してください。しかし、変換コネクションを追加したり削除したりするなど他の方法で、必要なときにはいつでもコンテンツ処理パイプラインを再構成することはできます。

変換コネクションを定義していなければ、パイプラインに変換コネクションを挿入することはできません。しかし、変換コネクションが定義してあり、コンテンツパイプラインに含めたいのであれば、変換コネクションプルダウンから選んで、説明を説明ボックスに書き、「前に変換を挿入」ボタンの1つをクリックしてコンテンツパイプラインに挿入します。

通知コネクションを定義していなければ、ジョブの終了時に通知を1つ以上追加することはできません。しかし、通知コネクションが定義してあり、含めたいのであれば、通知コネクションプルダウンから選んで、説明を説明ボックスに書き、「追加」ボタンをクリックして通知一覧に追加します。

ここでジョブの優先度及び開始方法を指定することができます。優先度とは、他ジョブと相対比較した場合にコンテンツを読み込む重要さです。高く設定された数値のジョブの方が最初に読み込まれます。開始方法とは先ほど説明したように、手動で開始、スケジュールされた日時に開始、他スケジュールされたジョブの後に開始です。

設定を指定した後に「次へ」ボタンを押下してください。その他のタブとページ下に「保存」ボタンが表示されます。ジョブを登録または更新する場合は必ず「保存」ボタンを押下してください。代わりに「キャンセル」ボタンを押せばコネクションは作成されません。(左に表示されるメニュー項目を選択しても同じことです。)

すべてのジョブには「スケジュール」タブがあります。スケジュールタブからは、スケジュール関連の設定を行なえます:



このタブでは、以下のパラメータの設定を行なえます:

- ジョブを継続的に実行するか、各コンテンツを一回だけ読み込むか
- コンテンツが無効になるまでの時間。無効になると索引から削除されます
- コンテンツの更新を確認する最小間隔
- コンテンツの更新を確認する最大間隔
- 初期コンテンツを再シーディングするまでの待ち時間

UIが示しているように、最後の4つのパラメータは継続実行の場合のみ有効です。

このページからスケジュール期間を定義することもできます。スケジュール期間とは、ジョブを実行することが可能な時間帯です。時間帯は開始日時(曜日、月、日、時、分)と最大実行時間(分)で指定します。各プルダウンメニューから複数の設定を選択された場合は、各プルダウンメニューで選択された設定の中の一つと一致した日時にジョブは開始されます。

スケジュールを入力した後に、「スケジュールの追加」ボタンを押下してください:



画面例ではジョブを土曜日と日曜日夜、午前2時から最大4時間（午前6時まで）に実行するように定義しています。

この他のタブは選択されたコネクションタイプによって異なります。これらのタブの詳細に付いて、選択された出力コネクション及びリポジトリコネクションの章を参照してください。

ジョブを保存すると、ジョブの設定の要約画面が表示されます。(どのコネクションを選んだかによって詳細は多少異なりますが)以下になります：

出力

変換コネクション一覧

出力コネクション一覧

権限

権限グループ一覧

ユーザーマッピングコネクション一覧

権限コネクション一覧

リポジトリ

リポジトリコネクション一覧

List Notification Connections

ジョブ

ジョブ一覧

状態とジョブ管理

状態レポート

コンテンツの状態

キューの状態

履歴レポート

履歴レポート

最大アクティビティ

最大バインド幅

結果履歴

その他

ヘルプ

ログアウト

ジョブ情報

名前: test

ステージ

タイプ

前掲

説明

接続名

ハイブリッド:

1.

リポジトリ

1.

Not変換

1

2.

変換

2.

Not出力

1

3.

出力

Notifications:

ステージ

Description

Connection name

No notification connections

優先順位:

5

開始メソッド:

自動的にスタートしない

スケジュールタイプ:

一度全てのドキュメントをスキャンする

最小再読み込み間隔:

適用可能ではありません

最大再クロール間隔:

適用可能ではありません

再シート間隔:

適用可能ではありません

スケジュールされた実行時間があります

リンクタイプの最大ネップ数 'child': 無制限

ネップ数モード:

入手できないドキュメントを削除します

1.

ルートパス

Convert path to URI?

ルール

http://xyz/index.html

No

含む/除外

ファイル/ディレクトリ

一致

include

file

\*

include

directory

\*

2.

3.

編集

削除

コピー

Reset seeding

下に「編集」、「削除」、「コピー」、「シーディングの再設定」の4つのボタンがあります。各ボタンの目的について順に述べます。

「編集」ボタンでは、戻ってジョブの詳細を編集できます。ジョブの詳細を変更したい場合にはこのボタンを使います。

「削除」ボタンでは、ジョブを削除できます。存在するジョブがもはや必要ない場合にこのボタンを使います。ManifoldCFでジョブを削除すると、そのジョブを使って索引を作成したコンテンツがすべて索引から削除されることに注意してください。

「コピー」ボタンでは、現在のジョブのコピーを編集できるようになります。大部分が現在のジョブの仕様に基づく新しいジョブを作成したい場合にこのボタンを使います。似たようなジョブを数多く作成する場合には便利でしょう。

「シーディングを再設定」ボタンでは、ManifoldCFからジョブのシーディングの履歴が消えます。シーディングとはコンテンツが追加されたり修正されたりしたことを発見した過程です。このボタンをクリックすると、ManifoldCFは次のクロールでリポジトリ内のコンテンツをすべて確認するようになります。これは頻繁に行なうべきことではありません。ManifoldCFは自身でこの情報を適切に管理し、ジョブの詳細が変更された時には必ず同じことを自動的に行ないます。コンテンツをすべて再確認しないといけなような方法でコネクタソフトウェアを更新した場合に、このオプションを使ってください。

### 1.9 ジョブの実行

ジョブの実行状態を把握するには、左メニューから「状態とジョブ管理」リンクを選択してください。以下のようなページが表示されます:



ジョブの現在の状態を表示するにはページ下の「更新」ボタンを押下してください。一番左の状態列にあるリンクを使って、ジョブを直接制御できます。可能な処理には次のようなものがあります：

- 開始(ジョブを開始)
- 最小限の開始(ジョブを開始しますが、できる限り作業を少なくします)
- 中断(ジョブを中断)
- 停止(ジョブを一時停止)
- 再開(ジョブを再開)
- 再実行(ジョブを中断して再度実行することと同じです)
- 最小限の再実行(ジョブを中断して、できる限り作業を少なくして再度実行することと同じです)

ジョブのキューにあるコンテンツに関しては、列「コンテンツ」、「処理中」、「処理済み」は特別な意味があります。「コンテンツ」列はジョブに属しているコンテンツすべての数です。「処理中」列はそのジョブで処理を待っているコンテンツの数です。「処理済み」列はジョブのキューに入っていて、一回以上は処理されたコンテンツの数です。

列举した処理のうち「最小限の」がついたものを使うと、そのジョブのコネクションタイプが使うモデルを仮定して最小限の作業量で動作します。ある場合には、これは追加や更新は索引が作られますが、削除は検出されないという意味になります。対象の索引とリポジトリのコンテンツを完全に同期するには、完全なジョブの実行が通常は必要です。



## 1.10 状態レポート

ManifoldCFのすべてのジョブはコンテンツ・セットに関連しています。セットに含まれるコンテンツの場所情報はジョブキューに保管されています。ManifoldCFのGUIページからこのキューを参照することができます。

各状態レポートから以下の情報でジョブキューのコンテンツを選択することができます：

- ジョブ
- コンテンツID
- コンテンツの状態と状況
- コンテンツが次に処理されるスケジュール

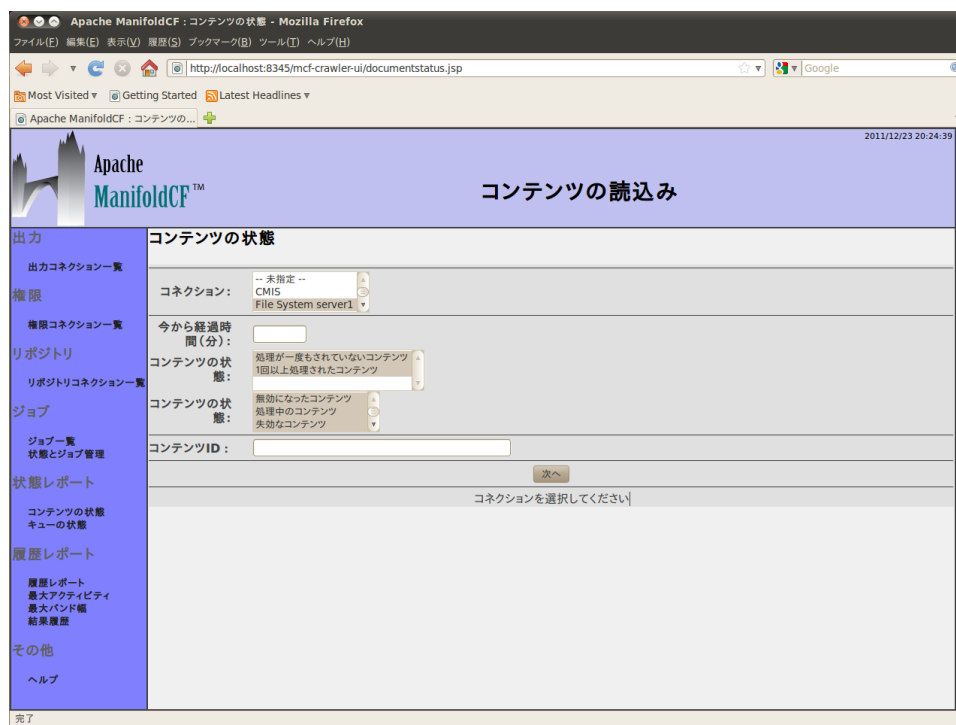
### 1.10.1 コンテンツ状態

コンテンツ状態レポートは、指定した条件に一致したコンテンツとその状態、状況、予定されている処理の一覧を表示します。実行中のジョブがコンテンツを処理したか確認する場合などに使うことができます。

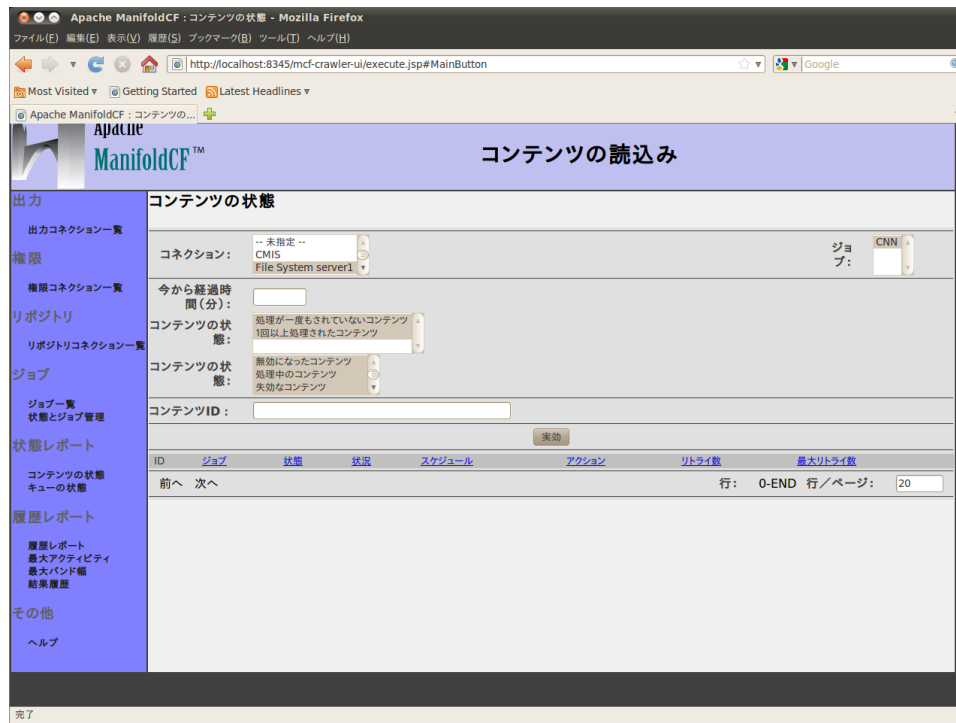
左メニューから「コンテンツ状態」リンクを選択すると、以下のようなページが表示されます：



検索するコネクションを選択してください。コンテンツの状態、状況と、コンテンツIDをフィルタする正規表現を指定することもできます。「次へ」ボタンを押下すると以下のようなページが表示されます：



右に表示されるドロップダウンリストから、ジョブを選択して、再び「次へ」ボタンを押下してください。以下のようなページが表示されます:



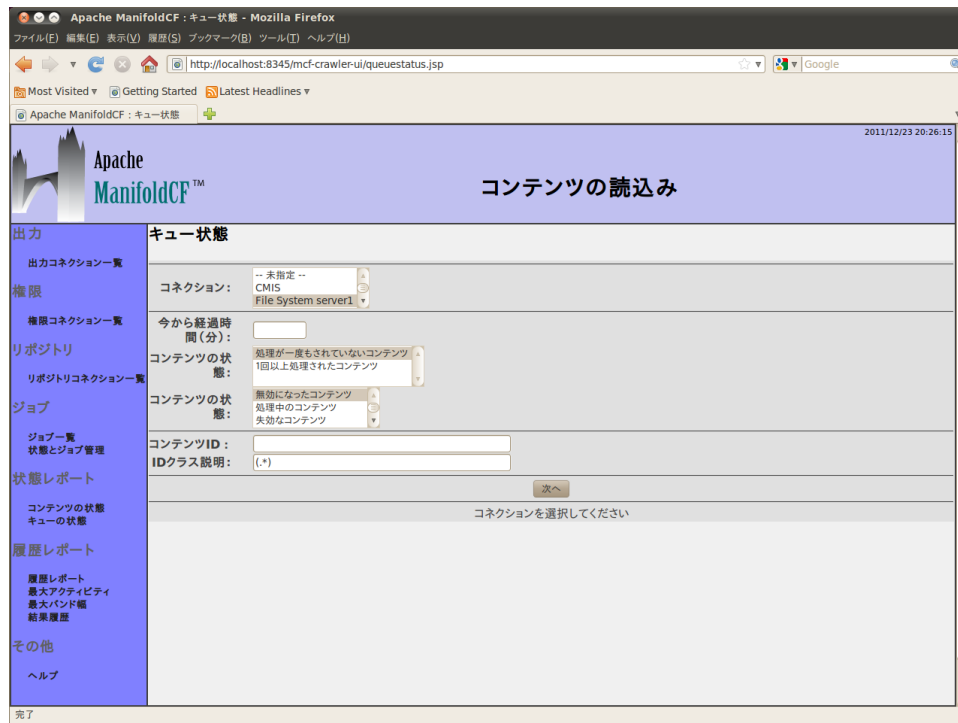
条件を変更して「実行」ボタンを押下して表示するコンテンツ情報を変更することもできます。また、表示する結果数を変更して「実行」ボタンを押下して、1ページに表示するコンテンツ数を変更することもできます。1ページにすべての一致したコンテンツが表示できない場合は、「前へ」リンクと「次へ」リンクを押下した表示する内容を移動することができます。

#### 1.10.2 キューの状態

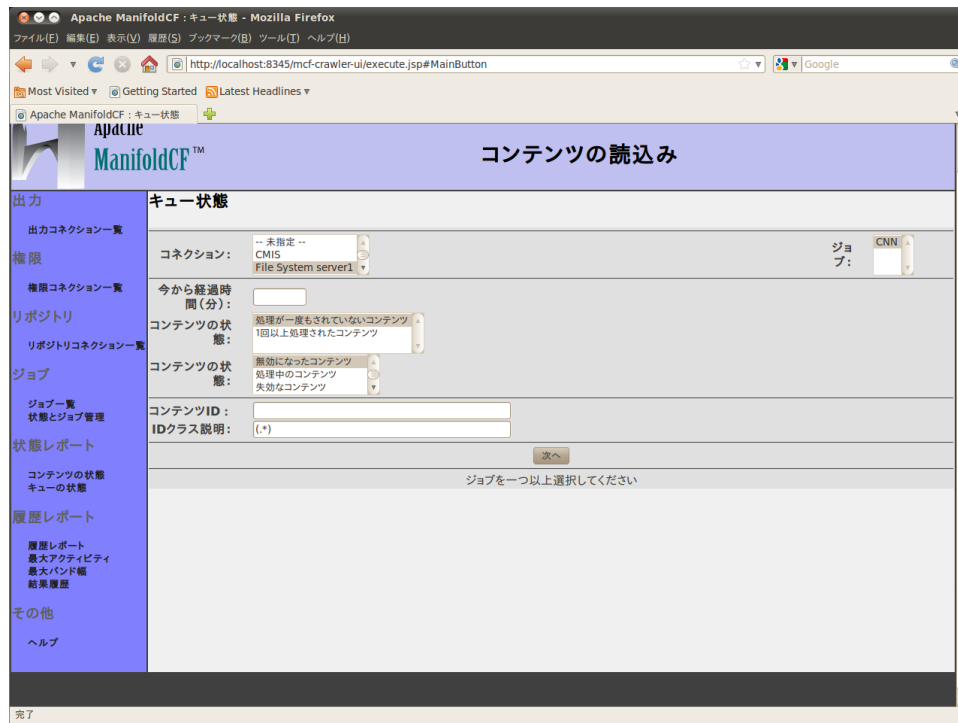
キューの状態レポートは、指定したclassに現れるコンテンツの回数を表示します。classは指定されたコンテンツIDに一致した正規表現のグループとして指定されます。結果はコンテンツの数として表示します。コンテンツの状態と状況の組み合わせ毎に列が設けられます。

例えば、「()」クラスを指定した場合は状態／状況の組み合わせ毎を1行で表示します。「(.\*)」クラスと指定した場合は、コンテンツID毎に行が設けられ、関連しているコンテンツの状態／状況の列に「1」が記入され、それ以外の列には「0」が記入されます。

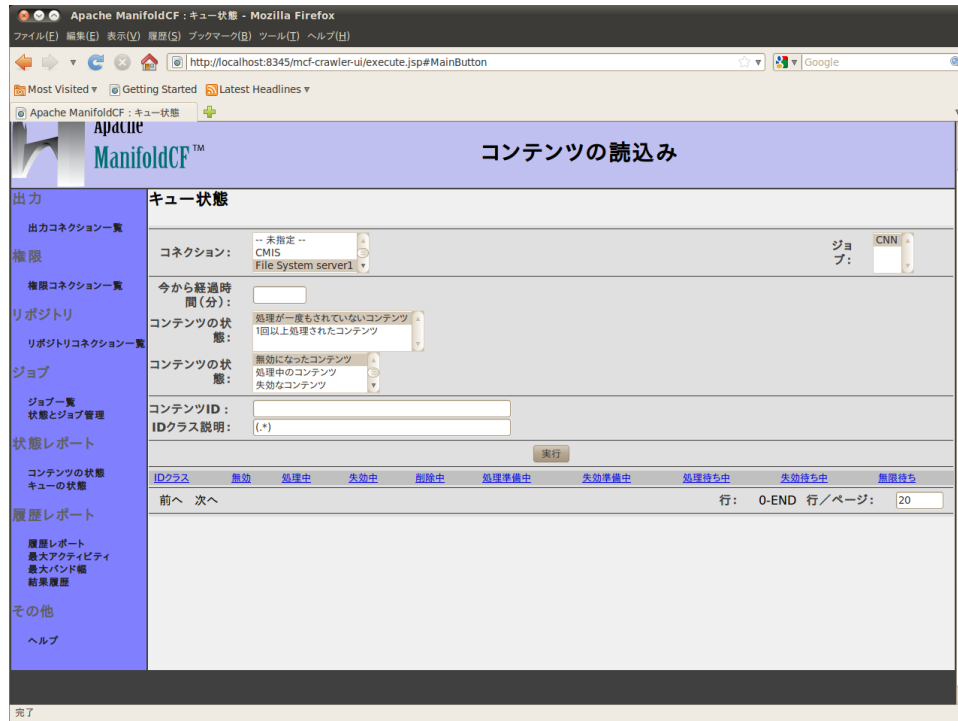
左メニューから「キューの状態」リンクを選択すると、以下のようなページが表示されます：



検索するコネクションを選択してください。コンテンツの状態、状況と、コンテンツIDをフィルタする正規表現を指定することもできます。コンテンツIDクラスはデフォルトでは「(.\*)」です。必要に応じて変更してください。「次へ」ボタンを押下すると以下のようなページが表示されます：



右の表示されるドロップダウンリストから、ジョブを選択して、再び「次へ」ボタンを押下してください。以下のようなページが表示されます：



条件を変更して「実行」ボタンを押下して表示するコンテンツ情報を変更することもできます。また、表示する結果数を変更して「実行」ボタンを押下して、1ページに表示するコンテンツ数を変更することもできます。1ページにすべての一致したコンテンツが表示できない場合は、「前へ」リンクと「次へ」リンクを押下した表示する内容を移動することができます。

### 1.11 履歴レポート

ManifoldCFは、コネクション毎にそのコネクションで起こった処理の履歴を記録しています。この履歴には、ManifoldCF基盤が記録したイベントと、リポジトリコネクション及び出力コネクションのイベントが含まれます。イベントは「アクティビティタイプ」として分類されます。以下のようなアクティビティタイプがあります：

- ジョブの開始
- ジョブの終了
- ジョブの中断
- 複数のconnection-type-specific読み込み及びアクセス処理
- 複数のconnection-type-specific出力及び索引作成処理

どのようにコンテンツを処理しているのかや、正しく動作しているのかを確認する場合に履歴レポートを使うことができます。ManifoldCFには履歴データを元にした複数のレポートが用意されています。

履歴レポートすべては、表示する内容を絞ることができるようになっています。以下の項目で絞り込む条件を指定できます：

- リポジトリコネクション名
- アクティビティタイプ (複数選択可)
- 開始時刻
- 終了時刻
- 対象とするコンテンツのID (正規表現で指定)
- 結果 (正規表現で指定)

レポートは処理問題や性能問題の原因を究明するのに使うことができます。各履歴レポートの詳細については以下の章を参照にしてください。

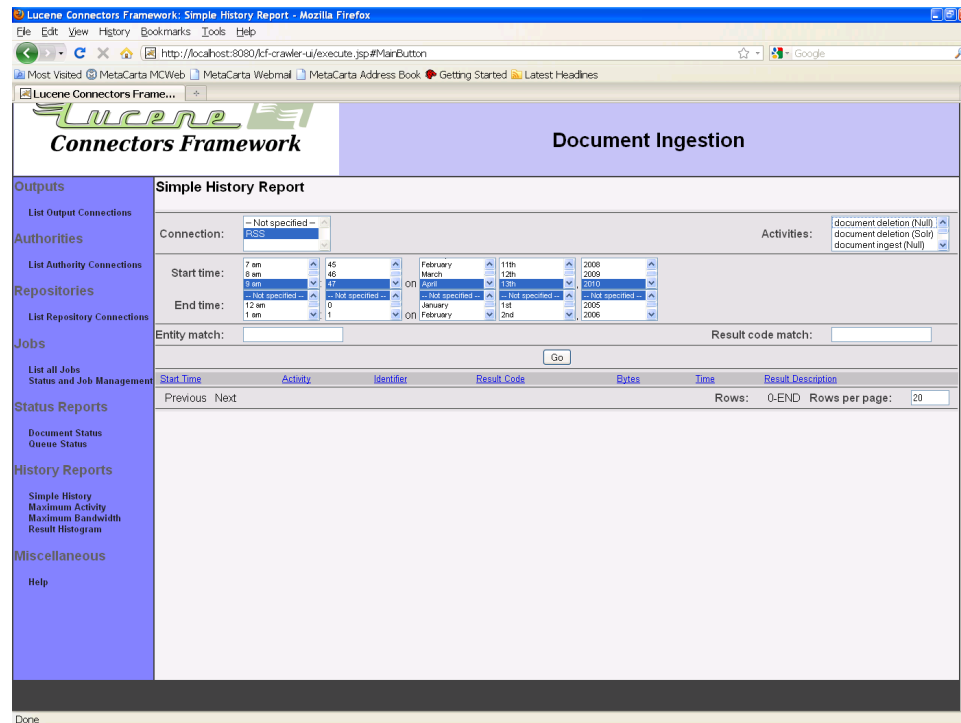
### 1.11.1 履歴レポート

履歴レポートは、集計などは行わずに、条件に一致したリポジトリコネクションの履歴データを表示します。最新イベントから古い順に開始時間、終了時間、処理、ID、データ量 (バイト)、結果などが表示されます。表示したレポートのイベント数を変えたり、指定した列順にソートしたり、ページを移動したりすることができます。

左メニューから「履歴レポート」リンクを選択してください。次のようなページが表示されます：

左上のプロダクションメニューからリポジトリコネクションを選択してください。開始と終了日付と時間、対象にするID及び結果コードを正規表現で指定することもできます。デフォルト設定では1時間前のすべてのイベントが選択されます。

「次へ」ボタンを押下してください。右上のプルダウンメニューからアクティビティが表示されます。



プロダクションメニューから1つ以上のアクティビティを選択し、「実行」ボタンを押下してください。該当するイベントが最新イベント順に表示します



違う条件で検索する場合は、条件を変更して「実行」ボタンを再び押下してください。また、表示する結果数を変更して「実行」ボタンを押下して、1ページに表示するコンテンツ数を変更することもできます。1ページにすべての一致したコンテンツが表示できない場合は、「前へ」リンクと「次へ」リンクを押下した表示する内容を移動することができます。

「実行」ボタンを押下すると、押された時間の条件での結果が表示されます。即ち、「実行」ボタンを押下した時点から1時間前までに起こったイベントが表示されます。ジョブが実行中の場合は、押す度に表示が変わる場合があります。

#### 1.11.2 最大アクティビティレポート

最大アクティビティレポートは、指定された時間帯に起こった最大のイベント発生率を表示します。(あとで追記)

#### 1.11.3 最大帯域幅レポート

最大帯域幅レポートは、指定された時間帯のイベントの最大バイト転送率を表示します。(あとで追記)

#### 1.11.4 結果ヒストグラムレポート

結果ヒストグラムレポートは、指定したイベントに一致する各結果の数を表示します。

## 1.12 認証についての覚え書き

選択されたコネクシオンタイプに認証が必要な場合は、システム管理者から必要な情報を入手してください。システム管理者は必要以上に力のあるアカウントと認証を提供するのをしばしば渋り、ときには提供すらしません。各コネクシオンは、コンテンツを読み取るのに最低限に必要とする認証で動作するように多大な注意を払って設計されています。もしコネクシオンの状態にセキュリティに関しての警告が表示された場合は、コネクシオンがその作業を完了するのに認証の権限が不十分であることをシステム管理者に知らせて、協力して問題を修正して下さい。

## 2 出力コネクシオンタイプ

### 2.1 Amazon Cloud Search出力コネクシオン

Amazon Cloud Search出力コネクシオンタイプは指定されたAmazon Cloud Searchインスタンス内の特定パスにドキュメントを送ります。このコネクシオンタイプは出来る限りコストを適切なおとところまで下げるため、更にドキュメントを一括処理します。従って、一部ドキュメントは通常のインデックシング時ではなくジョブ実行の最後に送信される場合があります。

Amazon Cloud Search出力コネクシオンタイプのコネクシオン設定情報には追加の「サーバー」タブが含まれています：

コネクシオンが機能するためには「サーバー名」を必ず入力しなければなりません。

ジョブ定義時にAmazon Cloud Search出力コネクシオンタイプと関連するタブは提供されません。

Amazon Cloud Search出力コネクシオンタイプはUTF-8と互換性のあるテキストコンテンツのみアクセス可能です。Amazon Cloud Search出力コネクシオンタイプに先立って、パイプライン経由でTika Content Extractorを用いてドキュメントをインデックシング可能な形式に変換することをお勧めします。

### 2.2 ElasticSearch出力コネクシオン

ElasticSearch出力コネクシオンは、XML文書をHTTP APIを介してElasticSearchに送ります。このコネクタは、できるだけ簡単に利用できるように設計されています。

ElasticSearch出力コネクシオンタイプを選択した後は、「引数」タブの項目をElasticSearchの設定に従って入力してください。各ElasticSearch出力コネクシオンは1つ

の索引と対応しています。一つ以上の索引を利用する場合は、索引毎に出力コネクションを作成してください。

引数:

- サーバアドレス: ElasticSearchインスタンスのURL。デフォルトURL (http://localhost:9200) はElasticSearchがManifoldCFと同じサーバで起動している場合のアドレスです。
- 索引名: コネクタは指定された索引にデータを作成します。

ジョブでElasticSearch出力コネクションを選択した場合は、「ElasticSearch」タブが表示されます。このタブから以下の設定を指定する事ができます:

- 対象とするコンテンツの最大サイズ(バイト)。デフォルトでは16MBです。
- 対象とするMIMEタイプ。リポジトリコネクションによっては正しく処理されません。
- 対象とするファイル拡張子。リポジトリコネクションによっては正しく処理されません。

すべてのアクティビティは履歴レポートから参照することができます。コネクタは3つのアクティビティに対応しています: コンテンツの読み込み(索引の作成)、コンテンツの削除、索引の最適化。ジョブが終了すると索引は最適化されます。

ElasticSearchの詳細についてはElasticSearchユーザマニュアルを参照してください。

## 2.3 ファイルシステム出力コネクション

ファイルシステム出力コネクションは、Unixユーティリティのwgetのようにローカルファイルシステムに文書を保管することができます。このコネクションタイプによって格納されたドキュメントは、メタデータまたはセキュリティ情報を含んでいませんが、バイナリ・ファイルのみから構成されています。

ファイルシステム出力コネクションタイプの接続構成情報には追加のタブを含みません。しかしながら、付加的なJobタブがあり、「出力パス」と呼びます。タブはこのように見えます。

ドキュメントを出力したいパスを入力して、「保存」をクリックしてください。

## 2.4 HDFS出力コネクション

HDFS出力コネクションは、UnixユーティリティのwgetのようにHDFS(Hadoop Distributed File System)に文書を保管することができます。このコネクションタイプによって格納されたドキュメントは、メタデータまたはセキュリティ情報を含んでいませんが、バイナリ・ファイルのみから構成されています。

HDFS出力コネクションタイプのための接続構成情報は「サーバー」タブという追加のタブを1つ含んでいます。このタブこのように見えます。

HDFSネームノードのURIおよびHDFSユーザー名を書き入れてください。両方とも必要となります。

HDFS出力接続タイプについては、「出力パス」と呼ばれる付加的なJobタブがあります。このタブこのように見えます。

ドキュメントを出力したいパスを入力して、「保存」をクリックしてください。

## 2.5 MetaCarta GTS出力コネクション

MetaCarta GTS出力コネクションタイプはHTTP APIを介してMetaCarta GTS検索エンジンにコンテンツを送ります。

GTSはHTML、XML、RTF、PDF、マイクロソフトオフィス文書のみ処理することができます。他型の文書から索引を作成することはできません。その制限により、大きな対象外のコンテンツは取得されません。

ジョブでGTSタイプ出力コネクションを選択すると、2つのタブが表示されます:「コレクション」と「コンテンツ・テンプレート」。この2つのタブからGTS特定機能を設定を行うことができます。

## 2.6 Null出力コネクション

null出力コネクションは、主にリポジトリコネクションタイプを開発する技術者向けに用意されています。実運用で使うことは少ないと思います。

Null出力コネクションタイプは索引及び削除リクエストをログするだけです。その他の処理は行いません。Null出力コネクション固有のタブはありません。

## 2.7 OpenSearchServer出力コネクション

OpenSearchServer出力コネクションは、XML文書をHTTP APIを介してOpenSearchServerに送ります。このコネクタは、できるだけ簡単に利用できるように設計されています。

OpenSearchServer出力コネクションタイプを選択した後は、「引数」タブの項目をOpenSearchServerの設定に従って入力してください。各OpenSearchServer出力コネクションは1つの索引と対応しています。一つ以上の索引を利用する場合は、索引毎に出力コネクションを作成してください。

引数:

- サーバアドレス: OpenSearchServerインスタンスのURL。デフォルトURL (http://localhost:8080) はOpenSearchServerがManifoldCFと同じサーバで起動している場合のアドレスです。
- 索引名: コネクタは指定された索引にデータを作成します。
- ユーザ名とAPIキー: OpenSearchServerインスタンスに接続するためのユーザ認証情報。ユーザが作成されていない場合は、空白にしてください。次の画像はOpenSearchServerで、認証情報が記載されている画面です。

ジョブでOpenSearchServer出力コネクションを選択した場合は、「OpenSearchServer」タブが表示されます。このタブから以下の設定を指定する事ができます:

- 対象とするコンテンツの最大サイズ(バイト)。デフォルトでは16MBです。
- 対象とするMIMEタイプ。リポジトリコネクションによっては正しく処理されません。
- 対象とするファイル拡張子。リポジトリコネクションによっては正しく処理されません。



Apache ManifoldCF™

# Document Ingestion

Sep 9, 2011 12:48:26 PM

Outputs

Authorities

Repositories

Jobs

Status Reports

History Reports

List Output Connections

List Authority Connections

List Repository Connections

List all Jobs  
Status and Job Management

Document Status  
Queue Status

Name Connection Scheduling Hop Filters OpenSearchServer Paths

Edit a Job

Max file size (bytes): 16777216


Allowed MIME Types :

application/msword  
application/vnd.ms-excel  
application/vnd.openxmlformats-officedocument.wordprocessingml.document  
application/vnd.openxmlformats-officedocument.spreadsheetml.sheet  
text/html  
application/pdf  
application/vnd.ms-powerpoint  
application/vnd.openxmlformats-officedocument.presentationml.presentation  
doc  
docx  
xls  
xlsx  
ppt  
pptx  
html  
pdf  
odt  
ods  
rtf  
txt

Allowed file extensions:

Save Cancel

すべてのアクティビティは履歴レポートから参照することができます。コネクタは3つのアクティビティに対応しています: コンテンツの読込み (索引の作成)、コンテンツの削除、索引の最適化。ジョブが終了すると索引は最適化されます。



# Document Ingestion

Sep 9, 2011 12:57:07 PM

**Outputs**

List Output Connections

**Authorities**

List Authority Connections

**Repositories**

List Repository Connections

**Jobs**

List all Jobs

Status and Job Management

**Status Reports**

Document Status

Queue Status

**History Reports**

Simple History Report

**Connection:** -- Not specified --  
Filesystem

**Activities:** Deletion (OpenSearchServer)  
Indexation (OpenSearchServer)  
Optimize (OpenSearchServer)

**Start time:** 9 am 55 July 7th 2009

**End time:** -- Not specified -- -- Not specified -- -- Not specified -- -- Not specified -- -- Not specified --

**Entity match:**  **Result code match:**

Go

Start Time	Activity	Identifier	Result Code	Bytes	Time	Result Description
09-09-2011 12:57:07.480	read document	/Users/keiler/Desktop/OSS - demotest/search-servlet/initial module	OK	0	1	
09-09-2011 12:57:10.140	Indexation (OpenSearchServer)	/Users/keiler/Desktop/OSS - demotest/search-servlet/bat	OK	1227	93	
09-09-2011 12:57:10.698	read document	/Users/keiler/Desktop/OSS - demotest/search-servlet/bat	OK	1227	188	
09-09-2011 12:57:06.521	Indexation (OpenSearchServer)	/Users/keiler/Desktop/browner - demo.txt	OK	251	289	
09-09-2011 12:57:06.371	Indexation (OpenSearchServer)	/Users/keiler/Desktop/OSS - demotest.xml	OK	193	61	
09-09-2011 12:57:06.347	read document	/Users/keiler/Desktop/OSS - demotest.xml	OK	193	144	
09-09-2011 12:57:05.996	Indexation (OpenSearchServer)	/Users/keiler/Desktop/opensearchserver-job-parameters.PNG	OK	128773	169	
09-09-2011 12:57:05.947	read document	/Users/keiler/Desktop/OSS - demotest/search-servlet/bat	OK	0	1	
09-09-2011 12:57:05.942	read document	/Users/keiler/Desktop/opensearchserver-job-parameters.PNG	OK	128773	275	
09-09-2011 12:57:05.834	Indexation (OpenSearchServer)	/Users/keiler/Desktop/opensearchserver-job-parameters.PNG	OK	508977	675	
09-09-2011 12:57:05.697	Indexation (OpenSearchServer)	/Users/keiler/Desktop/Demo OSS - Enterprise/screenshot.cap...	OK	508977	321	
09-09-2011 12:57:05.641	read document	/Users/keiler/Desktop/Demo OSS - Enterprise/screenshot.cap...	OK	508977	431	

OpenSearchServerの詳細についてはOpenSearchServerユーザマニュアルを参照してください。

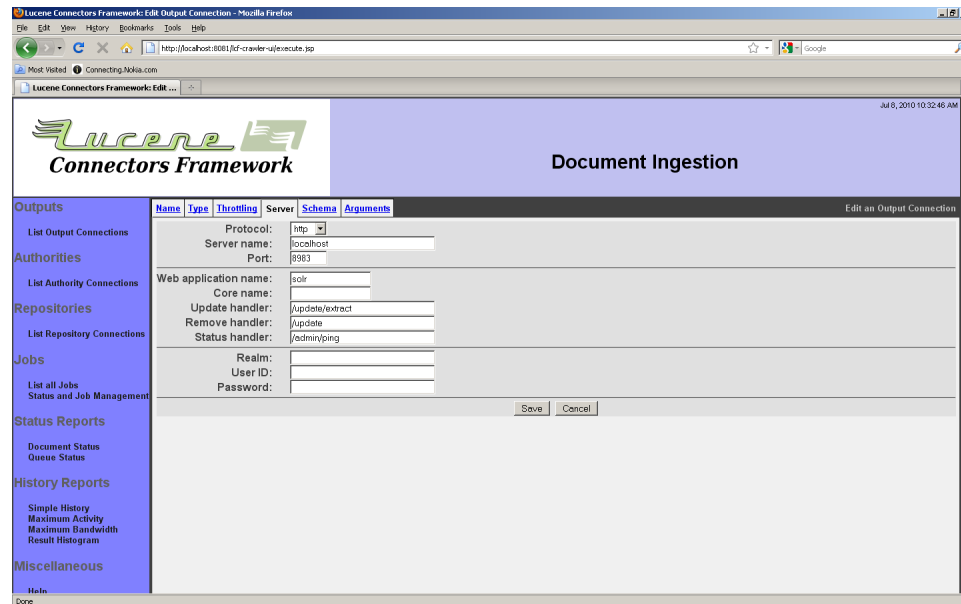
## 2.8 Solr出力コネクション

Solr出力コネクションタイプは、Solr HTTP APIを介してSolrにコンテンツを送ります。コネクションはSolrのデフォルト値にデフォルトで設定されます。Solrコネクションは索引可否に関係なく、すべてのコンテンツを処理します。設定されたパイプラインがコンテンツを利用するか判断するはずです。

ただし、すべてのコンテンツが送られるため映像のような大きなファイルもフィルタを設定しないと送られてしまい、システムに大きな負荷を掛けてしまいます。不足／間違っている設

定を発見してこのような問題を回避するために、Solrコネクションのすべてのクロール結果をレビューすることを推奨します。

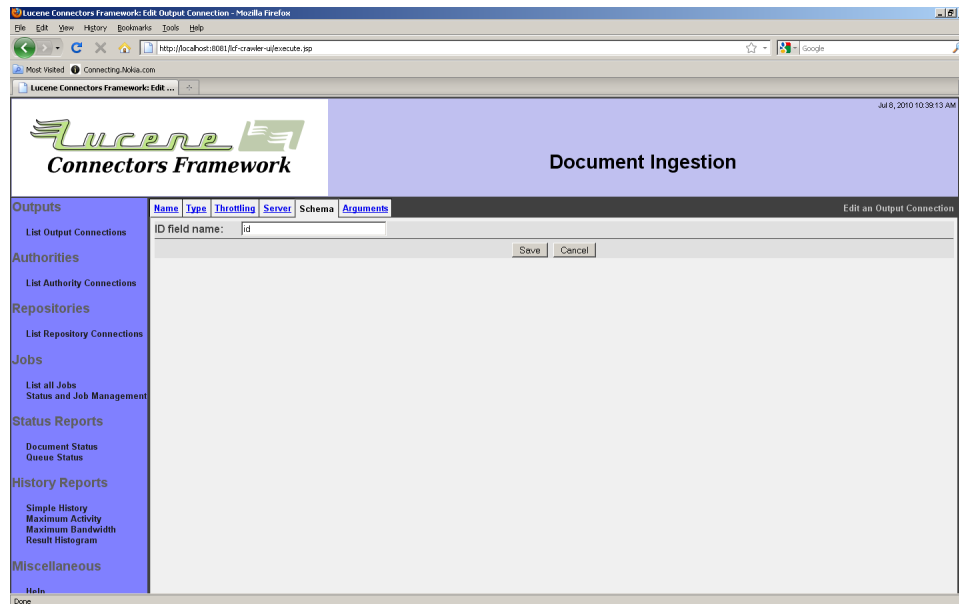
Solr出力コネクションを選択すると、5つのタブが表示されます。「サーバ」タブからHTTPターゲットを指定することができます：



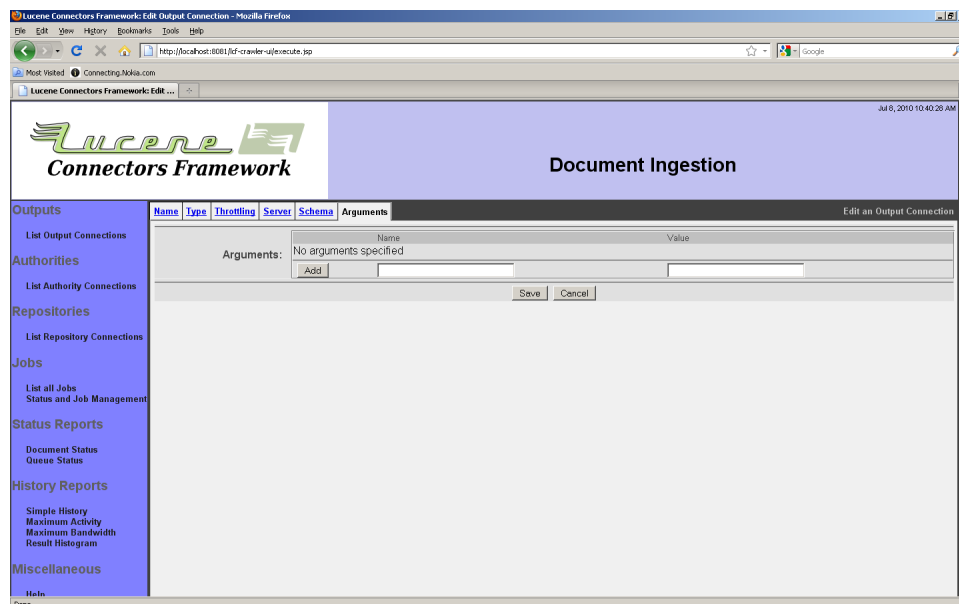
Solrの設定を入力してください。現リリースではベーシック認証のみ対応しています。下の部分にユーザ／パスワードを入力してください。

「スキーマ」タブからドキュメントIDに使うSolr項目を指定することができます。Solrコネクションはこの項目をコンテンツを検索するキーとして使います。





「引数」タブからはSolrに送る任意の引数を指定することができます。Solrの更新リクエストに利用できる任意の引数を利用することができます。たとえば、Solrのドキュメントを処理するために使われるパイプライン/チェーン: `update.chain=myChain`を追加することができます。その他に指定可能な引数についてはSolrのマニュアルを参照してください。タブは以下のように表示されます:

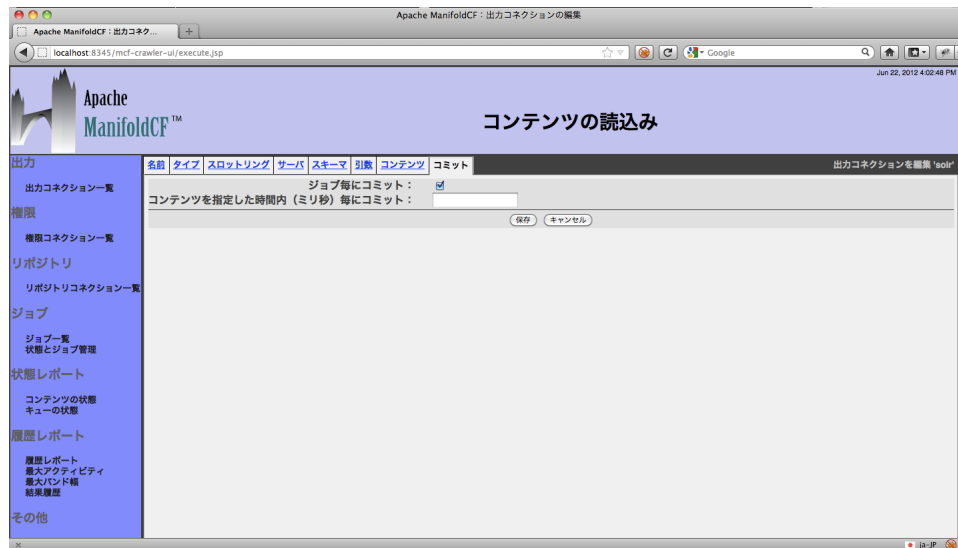


引数名と値を入力して、「追加」ボタンを押下してください。名前が既に存在する場合は、既存の値は新しく指定した値で置き換わります。引数を削除する場合は、削除する引数の左に表示されている「削除」ボタンを押下してください。

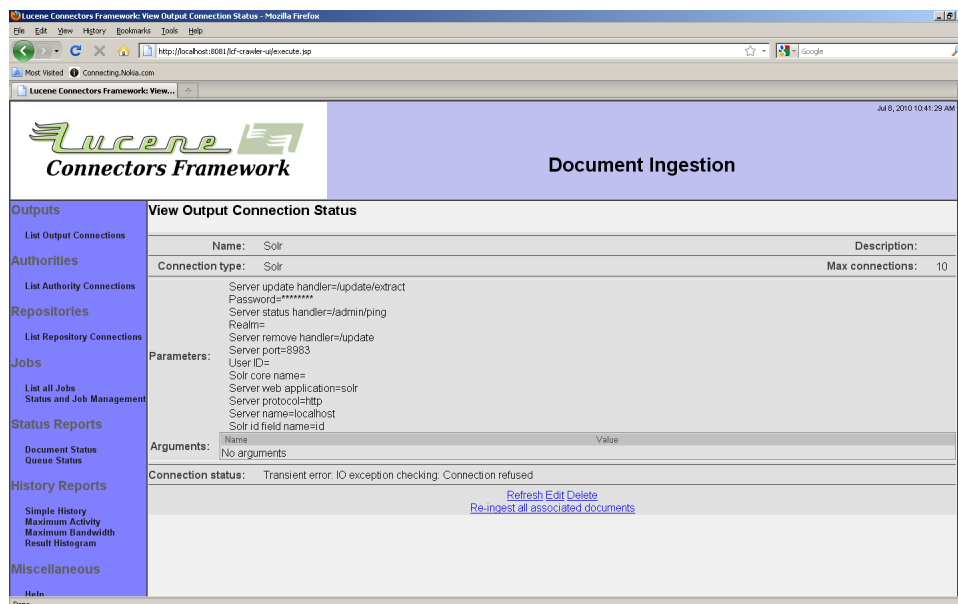
4番目のタブは"コンテンツ"タブです。これはドキュメントのサイズやMIMEタイプに基づいてフィルタリングを行うことができます。ドキュメントのバイト単位の最大長を指定することによって、そのサイズ（例えば10485760は10MBと同じです）を超えたドキュメントを除外することができます。特定のMIMEタイプのドキュメントだけを追加したい場合は、"含むMIMEタイプ"フィールドにそれらを入力することができます（例えばHTML以外のドキュメントを除外するなら"text/html"と登録します）。"除外するMIMEタイプ"フィールドは、特定のMIMEタイプのドキュメントを除外するためのものです（例えばJPEG画像を除外するなら"image/jpeg"と登録します）。タブは以下のように表示されます：



5番目のタブは"コミット"タブです。これはコミットの動作を制御することができます。すべてのジョブの終了時にドキュメントをコミットするようデフォルトで有効になっています。また、ミリ秒単位で一定時間内に各ドキュメントをコミットすることができます（10秒以内にコミットなら"10000"と登録します）。commit withinの挙動はManifoldCFでなくSolrに委ねられています。タブは以下のように表示されます：



設定の入力を完了した場合は、「保存」ボタンを押下してください。次のような入力した設定一覧が表示されます：



画面例では、Solrコネクションは正常に動作していないため、エラーメッセージが表示されています。

### 3 ユーザーマッピングコネクションタイプ

#### 3.1 正規表現ユーザーマッピングコネクション

正規表現ユーザーマッピングコネクションタイプは、あらゆる種類のユーザー名の機械的な変換に大変便利です。例えば、Active Directoryのユーザー名の標準的な"user@domain"形式からLiveLinkの"domain¥user"に変換するように簡単に設定できます。多くのリポジトリでそのような機械的な変換が確立されていますので、正規表現ユーザーマッピングコネクションタイプだけでたいの用は足ります。

正規表現ユーザーマッピングコネクションタイプを選択すると「ユーザーマッピング」タブが表示されます：

マッピングは一致する正規表現と置換文字列で構成されます。一致する正規表現は関心の部分を含む括弧(「(」と「)」)で括った正規表現のことです。括弧で括られた部分は正規表現の言葉で「グループ」と呼ばれます。置換文字列は固定文字と一致するグループの参照(変更されているかもしれませんが)から構成されます。例えば、「\$(1)」は1番目に一致したグループを参照し、「\$(11)」は小文字に変換された1番目に一致するグループを参照します。同様に、「\$(1u)」は同じ文字を参照しますが、大文字に変換されます。

例えば、一致正規表現`^(.*)¥@([A-Z|a-z|0-9|_|-]*\.)*$`と置換文字列`$(2)¥$(11)`はActive Directoryユーザー名MyUserName@subdomain.domain.comをユーザー名subdomain¥myusernameに変換します。

入力後に「保存」ボタンを押すと、次のようなコネクションの概要と状態のページが表示されます：

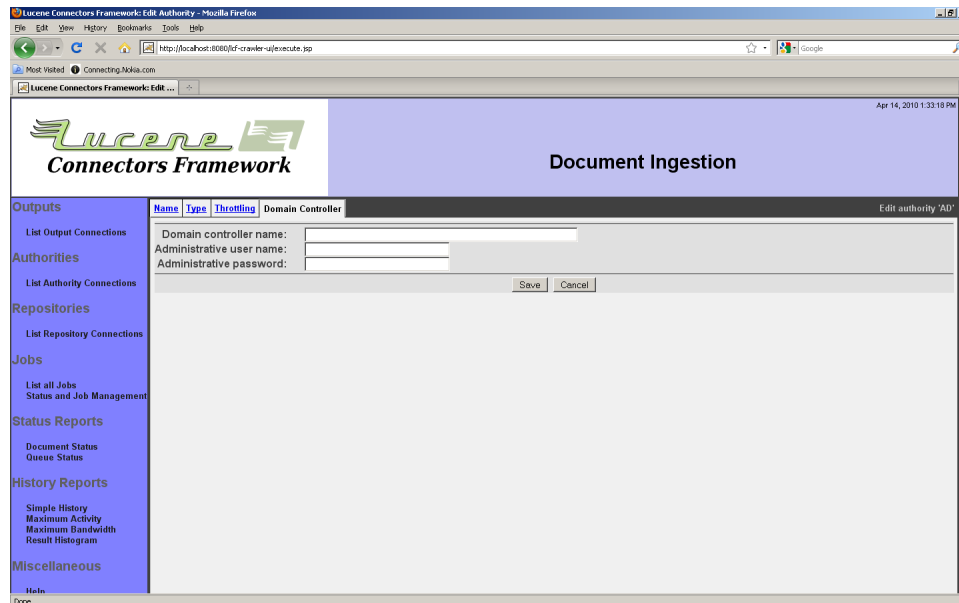
## 4 権限コネクションタイプ

### 4.1 Active Directory権限コネクション

Active Directory権限コネクションは、MS Windows共有ディレクトリ、(ActiveDirectoryモードでの)MS SharePoint、IBM FileNetリポジトリのファイル権限を有効に利用する場合に使います。Active Directory権限コネクションタイプを利用する場合は、Windowsドメインコントローラにログインして他ユーザIDとグループ関係を参照できる認証情報を設定する必要があります。

まだ以下のような場合では利用制限がありますが、一般的なWindowsセキュリティアーキテクチャを利用する場合に使うことができます：

Active Directory権限コネクションタイプは編集画面に独自タブが2つあります。「ドメインコントローラ」と「キャッシュ」です。「ドメインコントローラ」タブを選択すると次のようなページが表示されます：



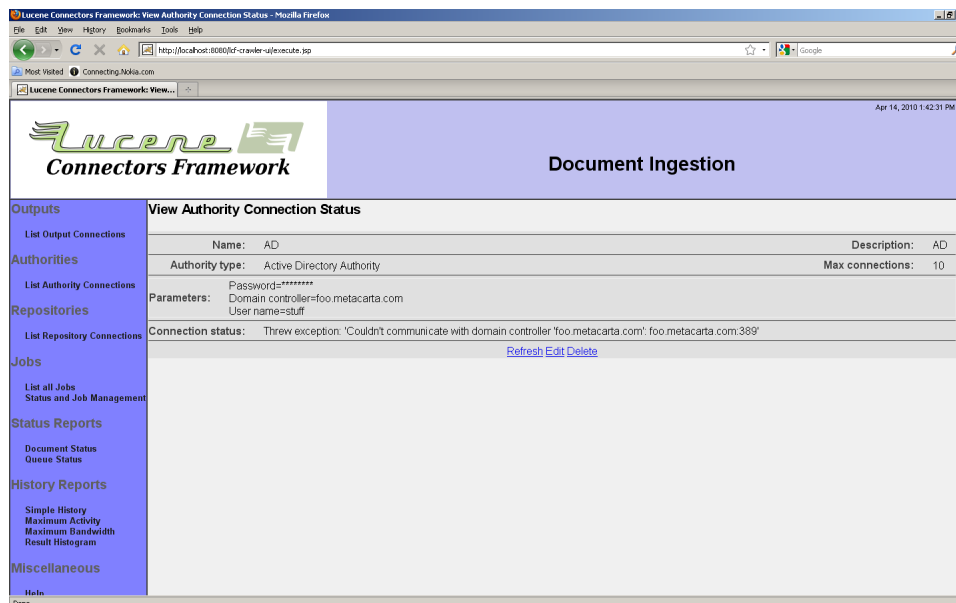
見ての通り、Active Directory権限では、異なっているがおそらく関連するドメインコントローラに対して複数のコネクションを設定できます。どのドメインコントローラにアクセスするかは、設定されたドメインコントローラの一覧を上から下まで見て、指定されたドメインサフィックスフィールドが最初に一致するものを選びます。ドメインサフィックスを空にするとすべてのユーザに一致すること注意してください。

一覧の最後にドメインコントローラを追加するには、必要な項目を入力してください。普通は「管理者ユーザ名」にはドメインを入力する必要はありませんが、ドメインコントローラの構成によっては「ユーザ名@ドメイン」形式で記入する必要があります。入力し終わったら「最後に追加する」ボタンを押して、一覧の最後にドメインコントローラを追加してください。後

で、他のドメインコントローラが一覧にあれば、適切な所にある別のボタンを押してドメインコントローラを好きな所に挿入できます。

Active Directory権限コネクションタイプには個々のユーザの応答のキャッシュを制御するための「キャッシュ」タブもあります。

ここではどれだけ多くの個々のユーザをどのくらい長くキャッシュするかを制御できます。入力した後に「保存」ボタンを押下すると、次のような設定概要と状態ページが表示されます：



画面ではActive Directoryと接続できないため、「Connection working」の代わりにエラーメッセージが表示されています。

## 4.2 LDAP権限コネクション

LDAP権限コネクションは、ネイティブなドキュメントセキュリティモデルがない状態でドキュメントセキュリティを提供するために使うことができます。例としては、Samba共有やWikiページやRSSフィード等が含まれます。

LDAP権限はLDAPサーバーからユーザーまたはグループ名をアクセストークンとして供給することで動作します。これらのアクセストークンは、ジョブごとに入力されたアクセストーク

ンを供給するリポジトリコネクション、またはSamba共有の、明示的なユーザ/グループ名をサポートしているJCIFSコネクションタイプで使うことができます。

このコネクションタイプは適切なLDAPサーバーにログインするための情報を入力する必要があります。サーチ表現もまたユーザやグループのレコードを検索するために必要です。この権限コネクションタイプはひとつの特殊なタブがあります。LDAPタブです：

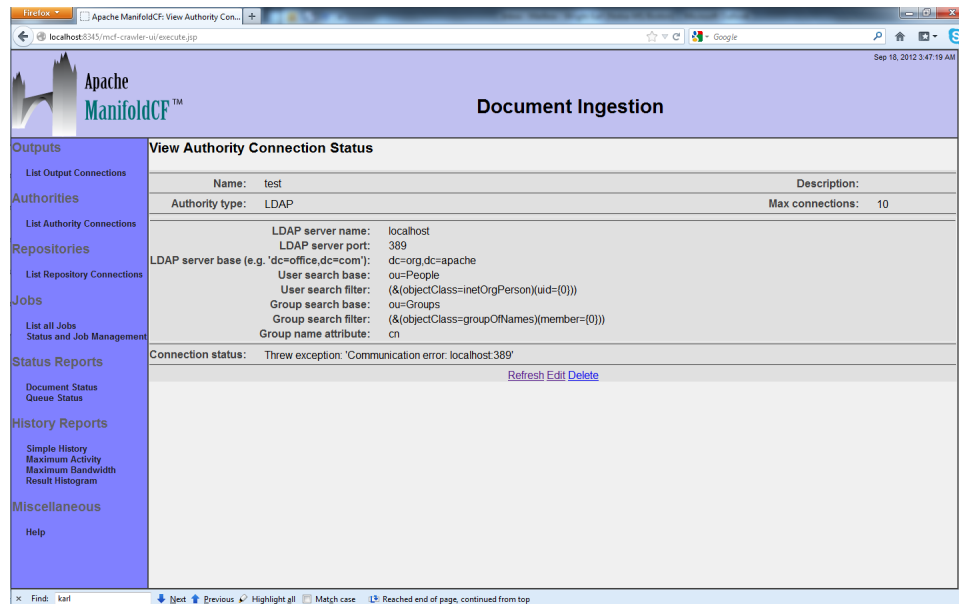
The screenshot shows the Apache ManifoldCF web interface in a Firefox browser. The page title is "Document Ingestion". On the left is a navigation menu with categories: Outputs, Authorities, Repositories, Jobs, Status Reports, History Reports, Miscellaneous, and Help. The "Authorities" section is active. The main content area shows the "LDAP" tab for configuring an authority. The form includes the following fields:

- LDAP server name: localhost
- LDAP server port: 389
- LDAP server base (e.g. 'dc=office,dc=com'): dc=org,dc=apache
- User search base: ou=People
- User search filter: (&(objectClass=inetOrgPerson)(uid=[\*]))
- Group search base: ou=Groups
- Group search filter: (&(objectClass=groupOfNames)(member=[\*]))
- Group name attribute: cn

At the bottom of the form are "Save" and "Cancel" buttons. The browser's status bar at the bottom shows "Find: karl" and navigation controls.

求められる値を入力してください。サーバーベースフィールドは検索したいLDAPドメインを含むことに注意してください。たとえばドメインがpeople.myorg.comならば、サーバーベースはdc=com,dc=myorg,dc=peopleとなります。

終わったらセーブボタンをクリックします。コネクションのサマリとステータスが表示されます。それは次のようなものになります：



注意点ですが、このサンプルではLDAPコネクションは応答していません。"Connection working"の代わりにエラーステータスのメッセージを表示しています。

#### 4.3 OpenText LiveLink権限コネクション

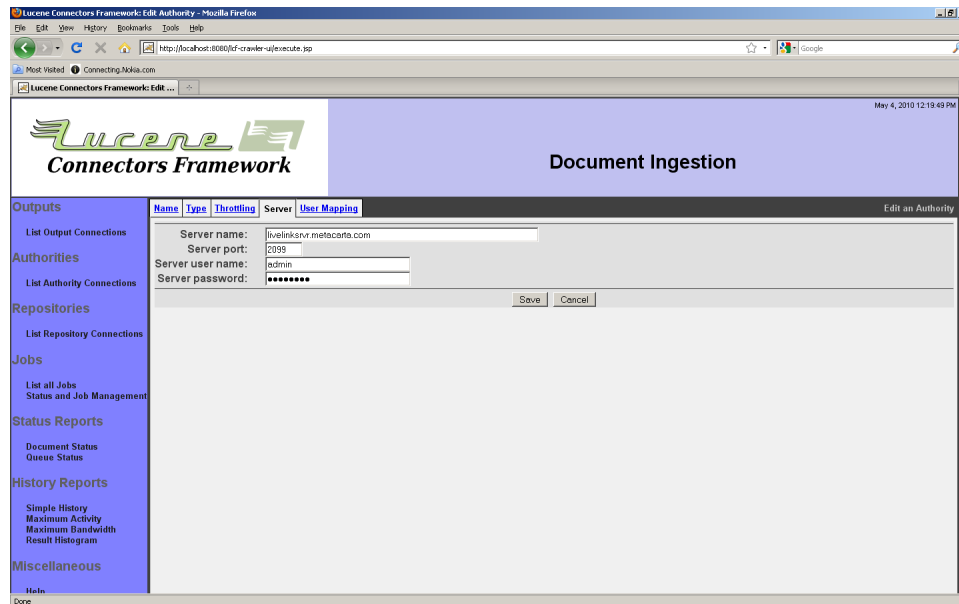
LiveLink権限コネクションは、LiveLinkリポジトリからコンテンツを取得する場合のセキュリティを指定する場合に利用します。

利用する場合はLiveLinkサーバ、ユーザのACLを取得できるユーザ認証情報を指定する必要があります。LiveLinkは独自でユーザ管理を行います。マッピングルールを定義または正規表現を利用してActive DirectoryとLiveLinkユーザと対応付けることもできます。

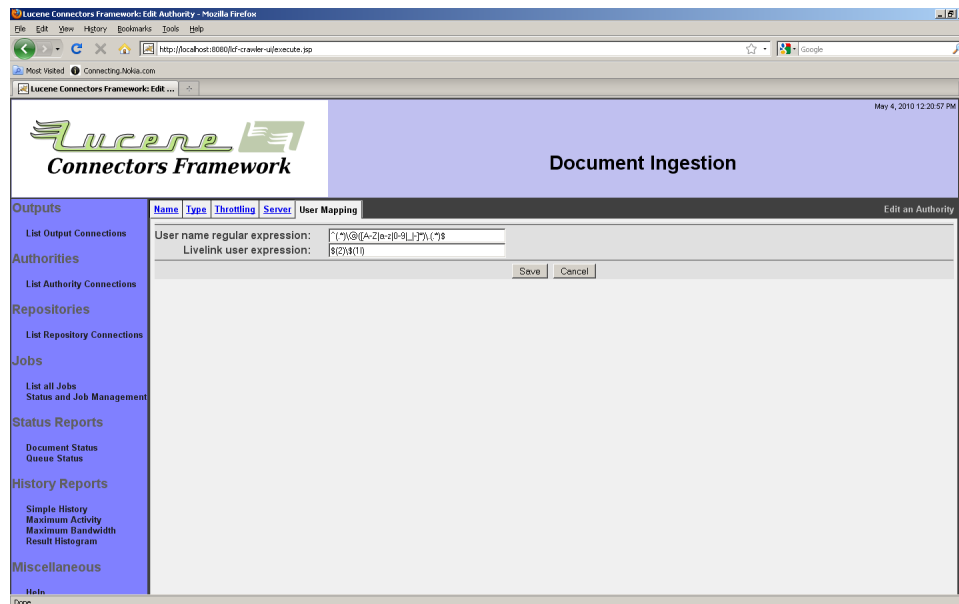
LiveLink権限コネクションを選択すると2つのタブが表示されます:「サーバ」タブと「ユーザマップ」タブ。

「サーバ」タブを選択すると以下のようなページが表示されます:





LiveLinkサーバ、ポート、認証情報を入力してください。  
「ユーザマップ」タブを選択すると次のようなページが表示されます：

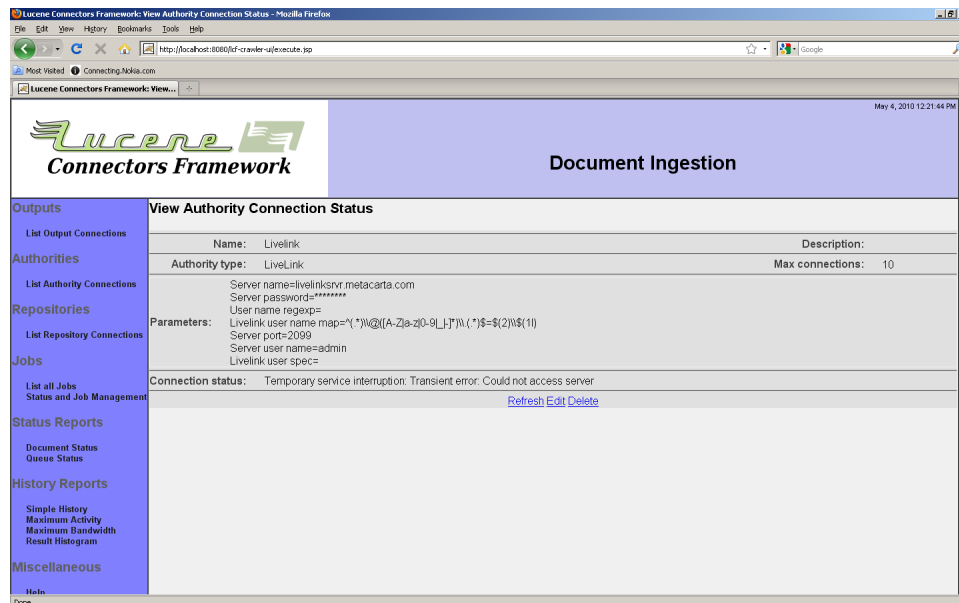


「ユーザマップ」タブから、ユーザ名及びドメイン（通常はActive Directoryから）からの情報をLiveLinkに対応付けることができます。対応は正規表現で定義します。変換元と値は括弧（「(」と「)」）で囲みます。括弧に囲まれた部分を「グループ」と言います。置き換え文字列は、固定文字と置き換えグループから構成されます。例えば、「\$(1)」は最初に一致したグルー

プを示し、「\$(1l)」は最初に一致した小文字のグループを示します。同じように「\$(1u)」は大文字にマップしたグループを示します。

例えば、一致条件`^(.*)¥@([A-Z|a-z|0-9|_|-]*)¥.(.*)$`と置き換え文字列`$(2)¥$(1l)`はActive Directoryユーザ名をMyUserName@subdomain.domain.comをLiveLinkユーザ名subdomain¥myusernameに対応付けます。

対応情報を入力した後に「保存」ボタンを押下すると、次のような概要及び状態情報が表示されます：



内容を確認してください。ページ例では、LiveLinkサーバに接続できないためエラーメッセージが表示されています。

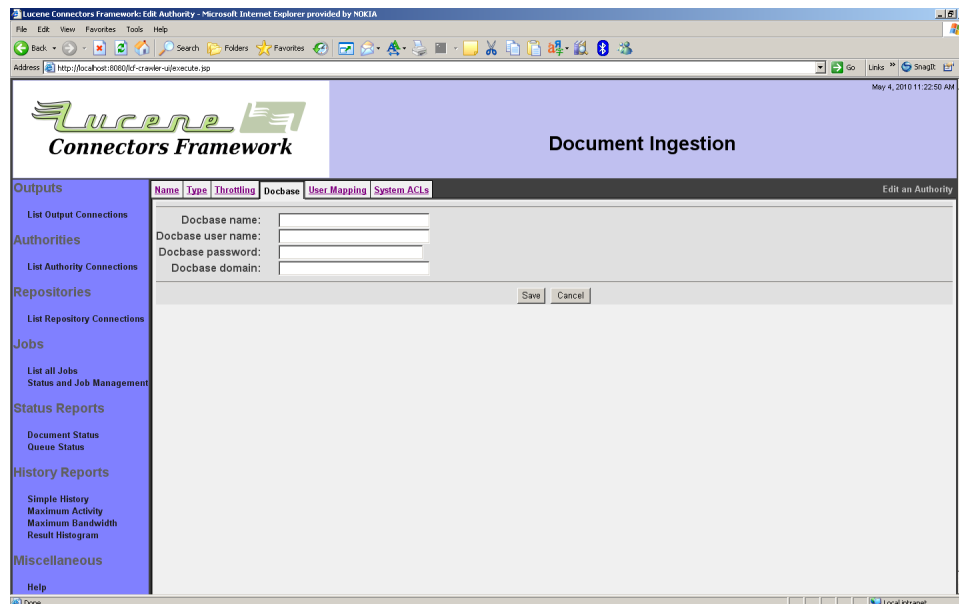
#### 4.4 EMC Documentum権限コネクション

Documentum権限コネクションは、Documentumリポジトリからコンテンツを取得する場合のセキュリティを指定する場合に利用します。

利用する場合は、Documentumコンテンツサーバの情報とユーザのACL情報を取得できる認証情報を指定する必要があります。ユーザー一覧毎に自動生成したACLを含むかも指定することができます。自動ACLはフォルダオブジェクト毎に生成されます。フォルダが多い場合は、ACLが多くなりユーザに戻されるManifoldCFアクセストークンが多くなり、性能を劣化させます。なお、多くの場合はDocumentumはこのACLを利用しません。そのため、多くの場合は、このACLを無視するように設定しても問題はありません。

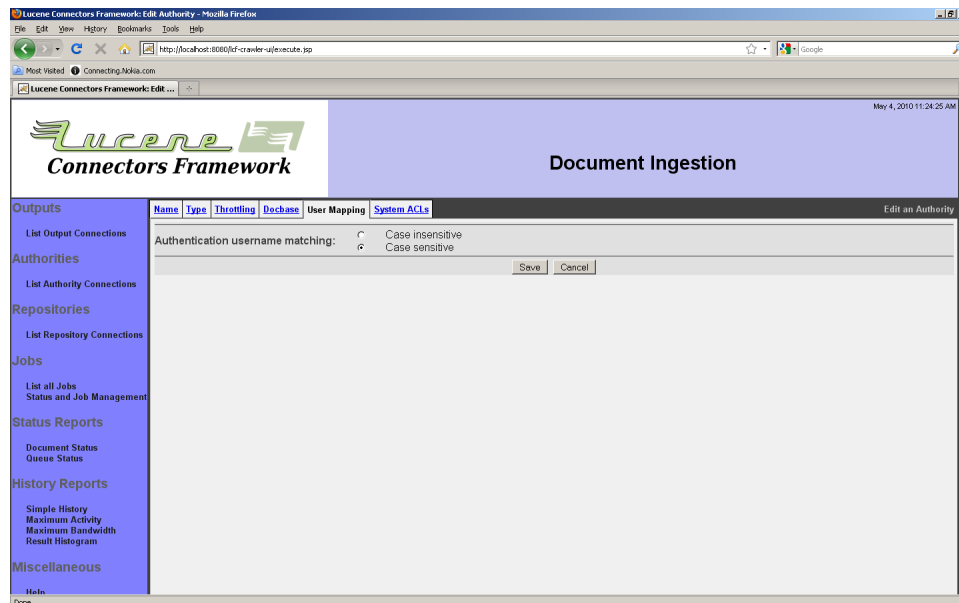
Documentum権限コネクションを選択すると3つのタブが表示されます：「Docbase」、「ユーザマップ」、「システムACL」。

「Docbase」タブを選択すると、次のようなページが表示されます：



コンテンツサーバdocbase名と認証情報を入力してください。コンテンツサーバでActive Directoryが有効に設定されていない場合は、項目「ドメイン」は空白にしてください。

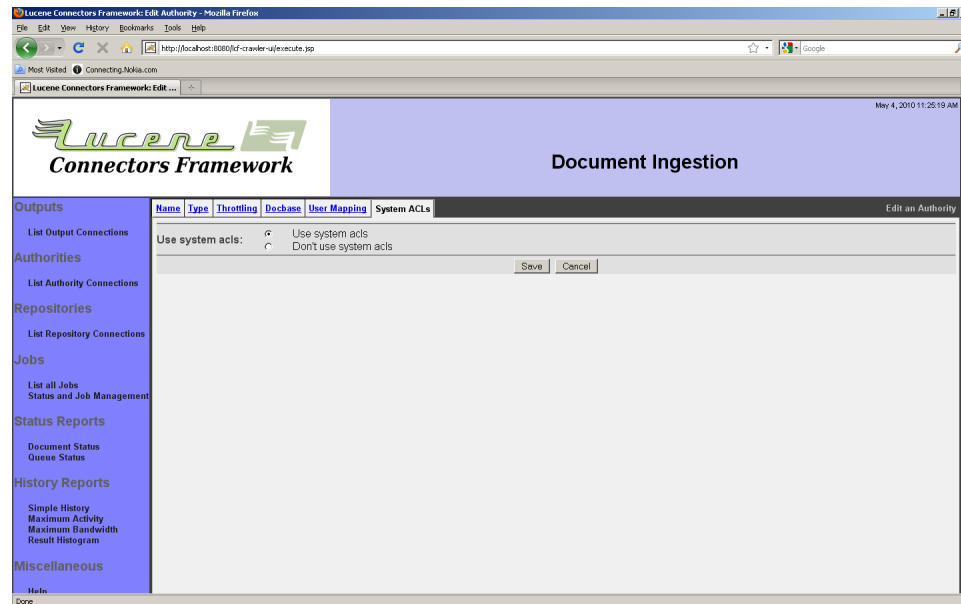
「ユーザマップ」タブを選択すると次のようなページが表示されます：



入力するユーザ名とコンテンツサーバユーザ名を対応付ける場合に大文字／小文字を区別するかを指定します。その他の対応は未対応です。多くの場合は、Documentumインス

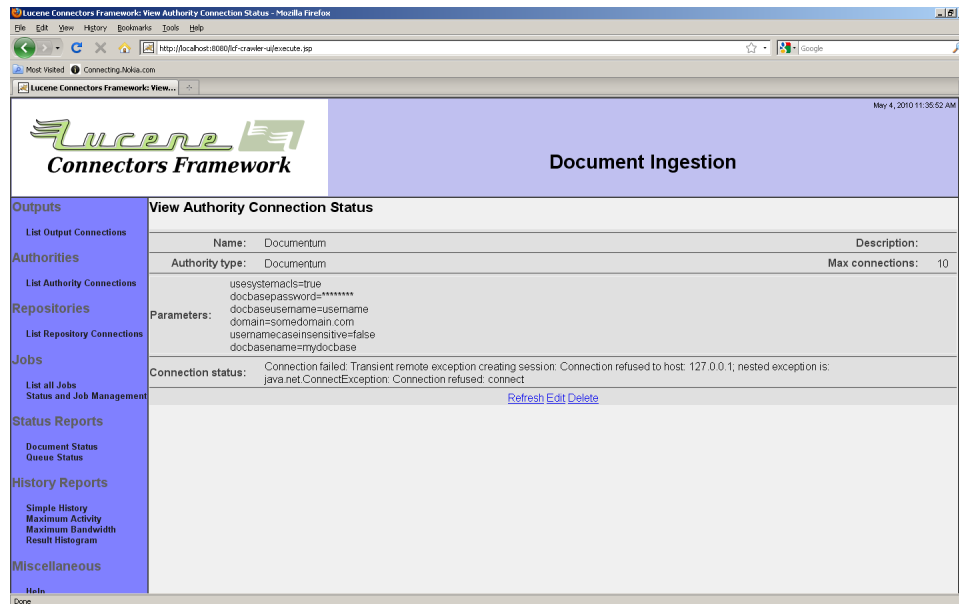
タンスはActive Directoryと連動して、Documentumユーザ名はActive Directoryのユーザ名と同じ、またはActive Directoryユーザ名はただ小文字にされています。詳細については、Documentumシステム管理者ガイドを参照してください。

「システムACL」タブを選択すると次のようなページが表示されます：



自動生成されたACLを無視するように指定することができます。まず無視するように設定して、必要であれば有効にすることを推奨します。Documentumシステム管理者に連絡して正しい設定を聞いてください。

入力した後に「保存」ボタンを押下すると、次のような概要及び状態情報が表示されます：



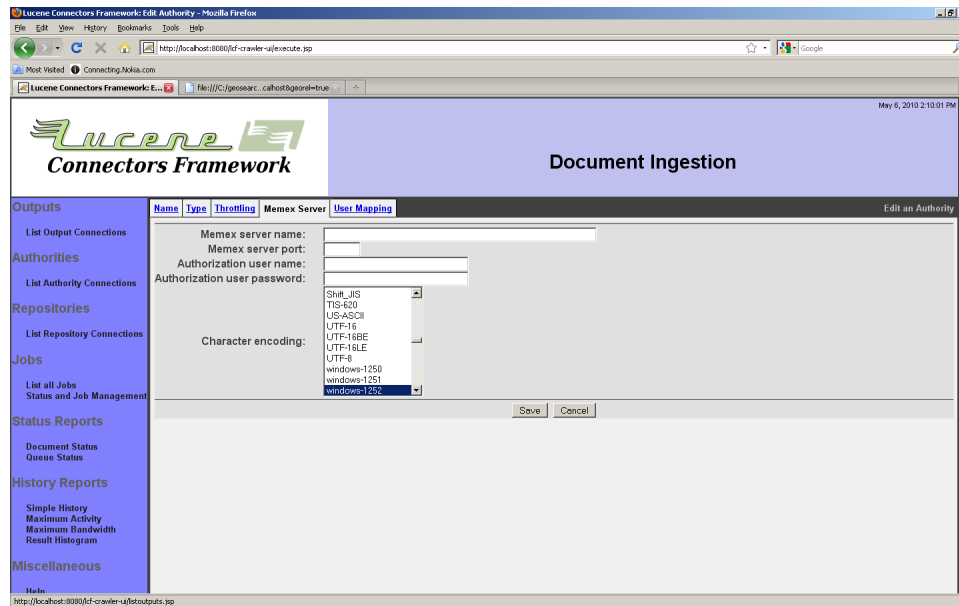
状態を確認して、必要であれば設定内容を修正してください。

#### 4.5 Memex Patriarch権限コネクション

Memex権限コネクションは、Memexリポジトリからコンテンツを取得する場合のセキュリティを指定する場合に利用します。

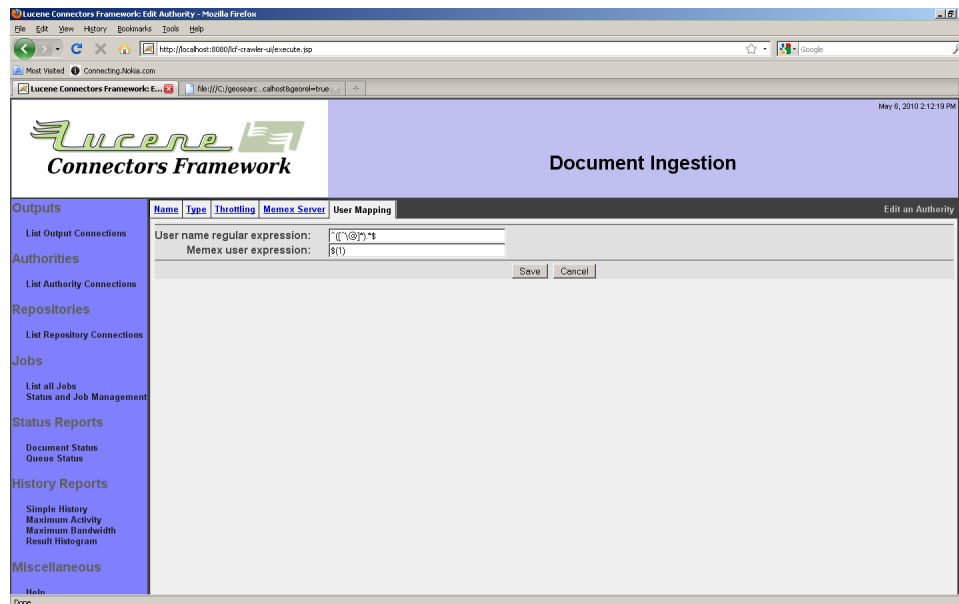
接続するMemexサーバとユーザマッピング情報、Memexサーバからユーザ権限情報を取得するためのユーザの認証情報を指定することが必要です。

Memex権限コネクションを選択すると、次の2つのタブが表示されます:「Memexサーバ」、「ユーザマップ」。「Memexサーバ」タブを選択すると次のようなページが表示されます:



Memexサーバ、ポート、Memexユーザ情報を取得できるユーザの認証情報を入力してください。また、Memexサーバも文字エンコーディングを選択してください。文字エンコーディングが不明な場合は、Memexシステム管理者に問い合わせてください。

「ユーザマッピング」タブを選択すると以下のようなページが表示されます：

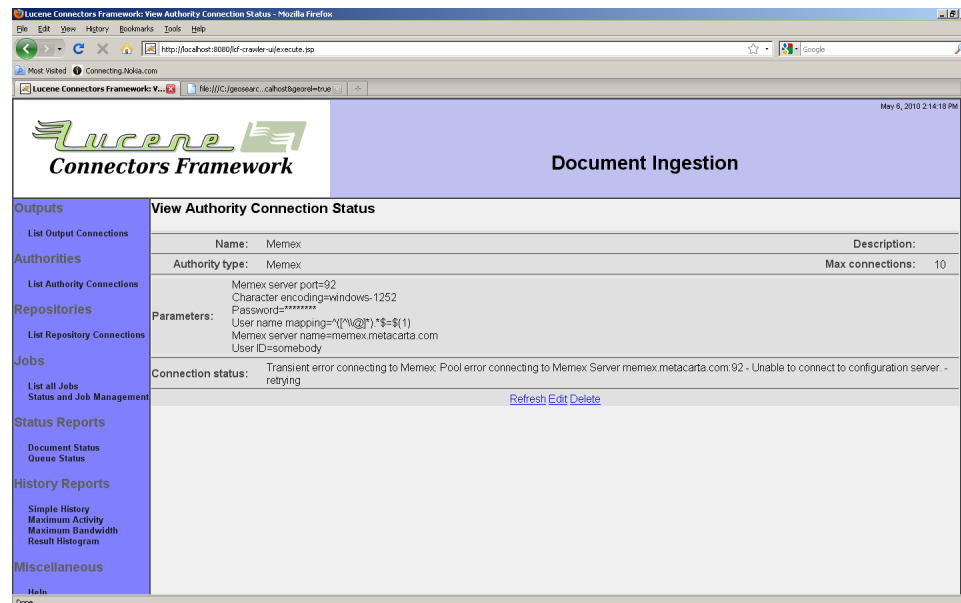


「ユーザマップ」タブから、ユーザ名及びドメイン（通常はActive Directoryから）からの情報をMemexに対応付けることができます。対応は正規表現で定義します 変換元と値は括弧（「(」と「)」）で囲みます。括弧に囲まれた部分を「グループ」と言います。置き換え文字列は、

固定文字と置き換えグループから構成されます。例えば、「\$(1)」は最初に一致したグループを示し、「\$(11)」は最初に一致した小文字のグループを示します。同じように「\$(1u)」は大文字にマップしたグループを示します。

例えば、一致条件`^([A-Z|a-z|0-9|_|-]*)¥.(*)$`と置き換え文字列`$(2)¥$(11)`は Active Directory ユーザ名を `MyUserName@subdomain.domain.com` を Memex ユーザ名 `subdomain¥myusername` に対応付けます

対応情報を入力した後に「保存」ボタンを押下すると、次のような概要及び状態情報が表示されます



内容を確認してください。ページ例では、Memexサーバに接続できないためエラーメッセージが表示されています。

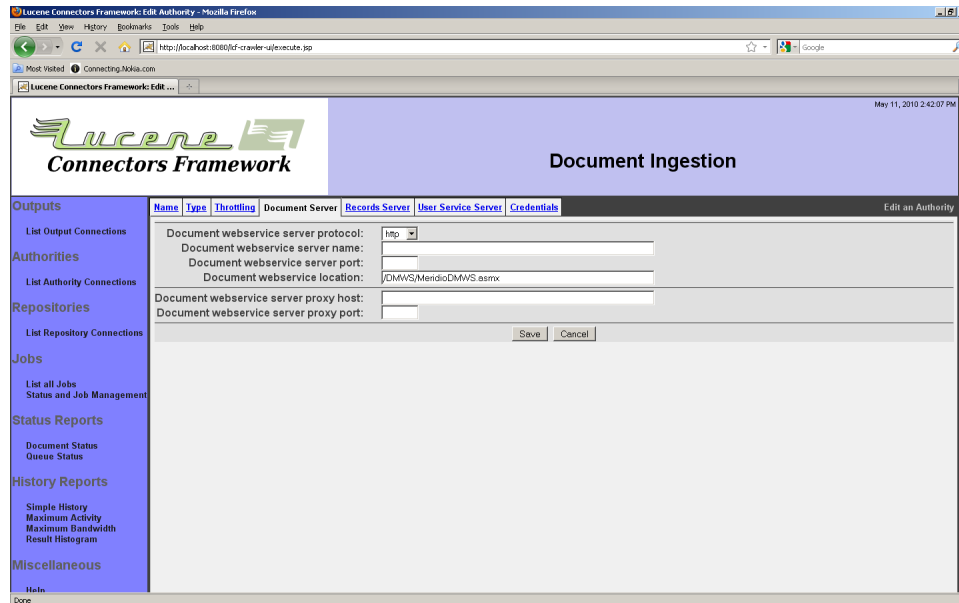
#### 4.6 Autonomy Meridio権限コネクション

Meridio権限コネクションは、Meridioリポジトリからコンテンツを取得する場合のセキュリティを指定する場合に利用します。

接続するドキュメントサーバ、レコードサーバ、ユーザサービスの情報を指定してください。ユーザのACL情報を取得するために利用するMeridioユーザの認証情報も必要です。

ユーザサービスはMeridio Authorityの一部です。Meridio Authorityを利用する場合は、Meridioシステムにインストールしてください。不明な場合は、Meridioサーバ管理者に問い合わせてください。

Meridio権限コネクションを選択すると4つのタブが表示されます:「ドキュメントサーバ」、「レコードサーバ」、「ユーザーサービスサーバ」、「認証」。「ドキュメントサーバ」タブを選択すると次のようなページが表示されます:

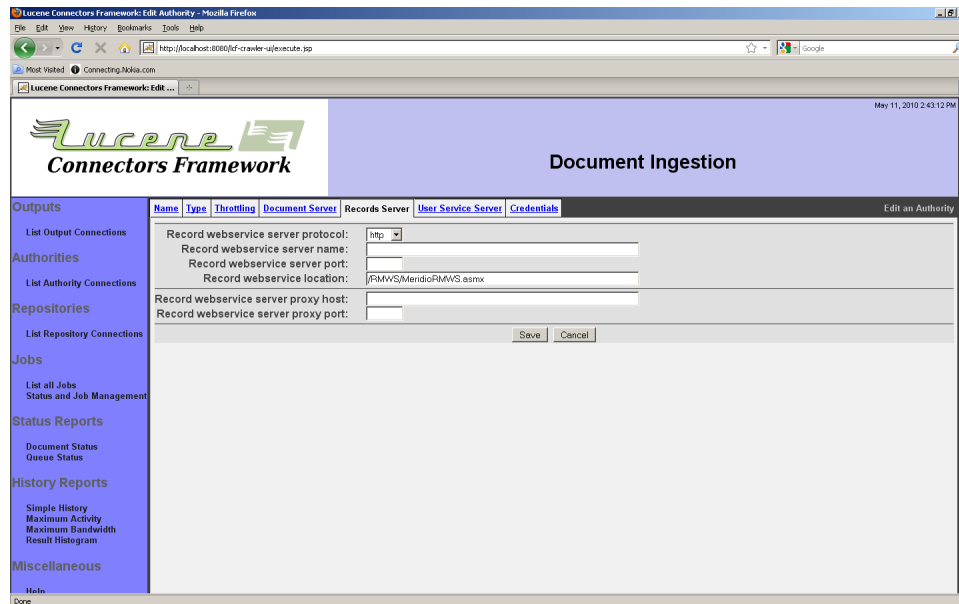


プロトコル、サーバ名、ポート、Meridioドキュメントサーバサービスのアドレスを入力してください。プロキシを利用している場合は、プロキシホストとポート番号も入力してください。認証プロキシは現リリースでは未対応です。

Meridioシステムの場合は異なるサービス毎にサーバを設けることができますが、一般には複数のサービスが同じサーバで動作しています。ただし、コネクションタイプ設定からは異なるサーバを指定することもできます。

「レコードサーバ」タブを選択すると、次のようなページが表示されます:

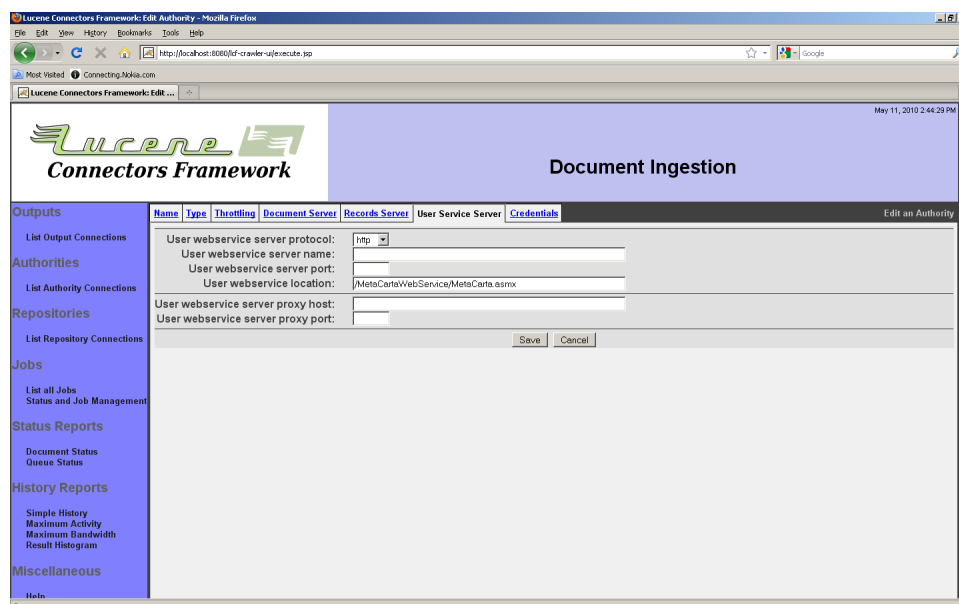




プロトコル、サーバ名、ポート番号、Meridioレコードサーバサービスのアドレスを入力してください。プロキシを利用している場合は、プロキシホストとポート番号も入力してください。認証プロキシは現リリースでは未対応です。

Meridioシステムの場合は異なるサービス毎にサーバを設けることができますが、一般には複数のサービスが同じサーバで動作しています。ただし、コネクションタイプ設定からは異なるサーバを指定することもできます。

「ユーザーサービスサーバ」タブを選択すると次のようなページが表示されます：

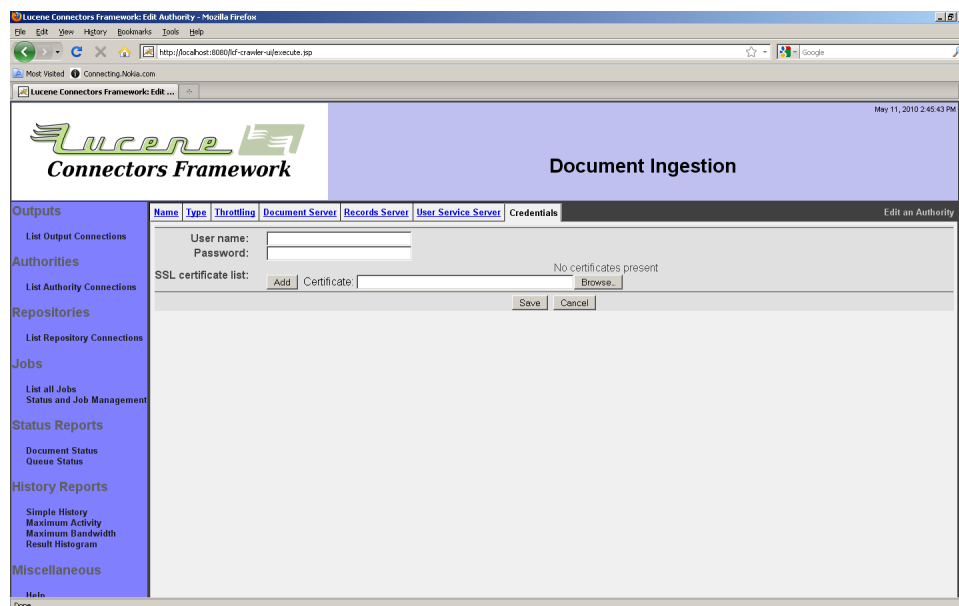


ページ内容を記入するには、Meridio Authorith extensionのインストール先を知る必要があります。

プロトコル、サーバ名、ポート番号、Meridioユーザサービスサーバサービスのアドレスを入力してください。プロキシを利用している場合は、プロキシホストとポート番号も入力してください。認証プロキシは現リリースでは未対応です。

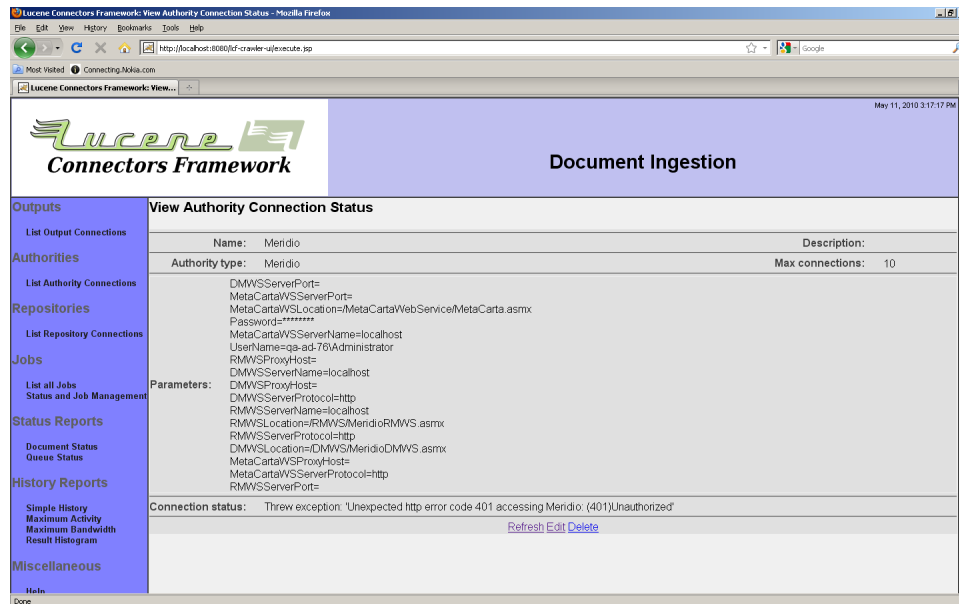
Meridioシステムの場合は異なるサービス毎にサーバを設けることができますが、一般には複数のサービスが同じサーバで動作しています。ただし、コネクションタイプ設定からは異なるサーバを指定することもできます。

「認証」タブを選択すると、以下のようなページが表示されます：



Meridioシステム用ユーザの認証情報を入力してください。

入力したら、「保存」ボタンを押下してください。次のようなページが表示されます：



表示されている画面ではMeridio権限サーバがWindowsドメインに接続できないためエラーになっています。

MeridioはWindows IISの認証機能を利用します。IIS及びWindowsドメインが正しく設定されていない場合は、Meridioも正常に動作しない場合があります。問題が発生した場合は、Meridio担当技術者に問い合わせてください。また、以下のようなデバッグツールを使うこともできます：

- Windowsセキュリティイベントログ
- ManifoldCFログ(以下の参照)
- パケットキャプチャ(例:WireShark)

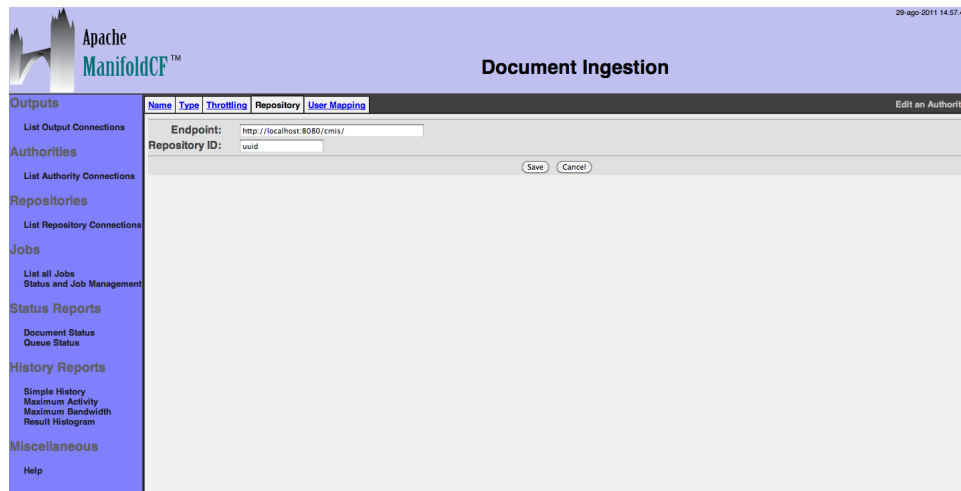
特定のManifoldCFログ情報が必要な場合は、ManifoldCF担当者に連絡してください。

#### 4.7 CMIS権限コネクション

CMIS権限コネクションは、CMISリポジトリから取得する場合のセキュリティを指定する場合に利用します。

CMIS仕様で特定のコンテンツに関しての権限設定をできるようになっている場合は、正規表現で指定することができます。

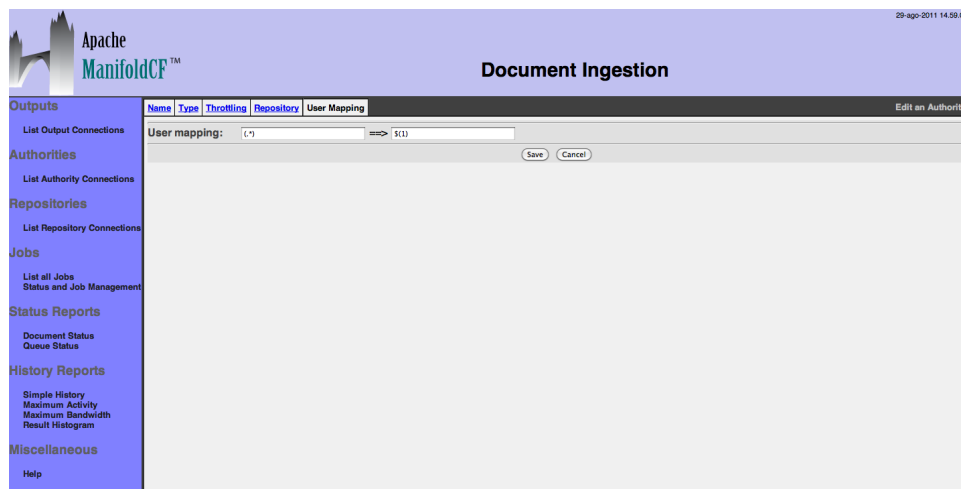
CMIS権限コネクションを選択すると、次の2つのタブが表示されます：「リポジトリ」、「ユーザマップ」。「リポジトリ」タブを選択すると次のようなページが表示されます：



リポジトリ設定は特定のCMISリポジトリのIDを追跡するためだけに使われます。CMISリポジトリを検索しません。

「ユーザマップ」タブからユーザの対応付けを指定することができます。

「ユーザマップ」タブを選択すると次のようなページが表示されます：



「ユーザマップ」タブから、ユーザ名及びドメイン（通常はActive Directoryから）からの情報をCMISに対応付けることができます。対応は正規表現で定義します。変換元と値は括弧（「(」と「)」）で囲みます。括弧に囲まれた部分を「グループ」と言います。置き換え文字列は、固定文字と置き換えグループから構成されます。例えば、「\$(1)」は最初に一致したグループを示し、「\$(11)」は最初に一致した小文字のグループを示します。同じように「\$(1u)」は大文字にマップしたグループを示します。

例えば、一致条件`^(.*)¥@([A-Z|a-z|0-9|_|-]*)¥.(.*)$`と置き換え文字列`$(2)¥$(1)`は Active Directory ユーザ名を `MyUserName@subdomain.domain.com` を CMIS ユーザ名 `subdomain¥myusername` に対応付けます。

対応情報を入力した後に「保存」ボタンを押下すると、次のような概要及び状態情報が表示されます：

The screenshot shows the Apache ManifoldCF web interface. The top header is 'Document Ingestion' with a timestamp '29-ago-2011 15:15:45'. The left sidebar has a blue background with white text for navigation: Outputs, Authorities, Repositories, Jobs, Status Reports, History Reports, and Miscellaneous. The main content area is titled 'View Authority Connection Status'. It shows a table with the following data:

Name:	Description:
CMIS Authority	
Authority type:	CMIS
Parameters:	usermapping=(.*)¥\$(1) repositoryId=uuid endpoint=http://localhost:8080/cmisis/
Connection status:	Connection working

At the bottom of the table, there are links: Refresh Edit Delete.

## 5 リポジトリコネクションタイプ

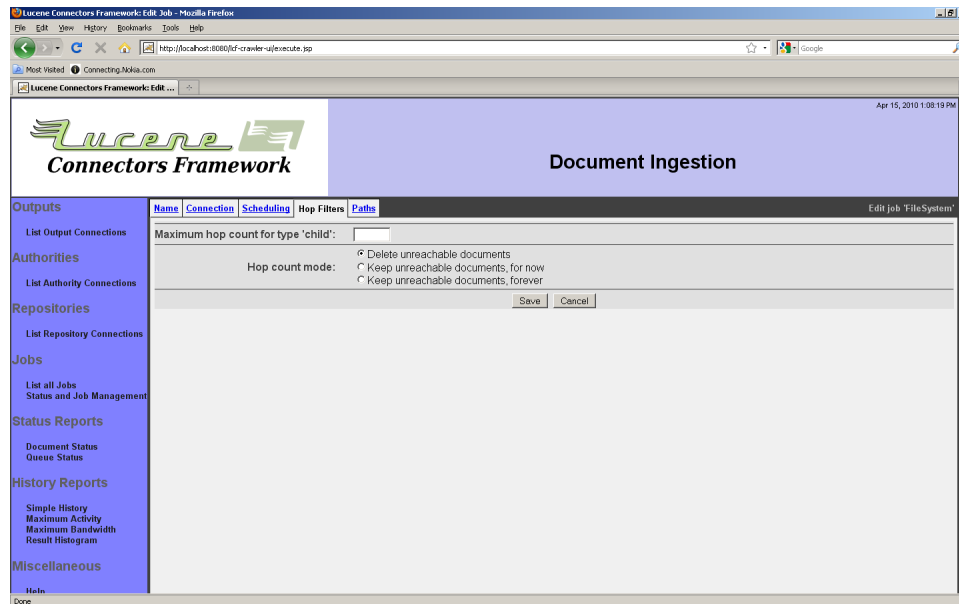
### 5.1 汎用ファイルシステムリポジトリコネクション

汎用ファイルシステムリポジトリコネクションタイプは主に例題、デモ、テストツールのために開発されました。ManifoldCFがインストールされているサーバのローカルファイルの索引を作成します。ただし、権限設定を行うことはできません。

ファイルシステムリポジトリコネクションタイプ固有のタブはありません。ただし、性能のために「スロットリング」タブの「最大接続/JVM」値をワーカスレッド毎に最低でも1つ、または30に設定してください。

ジョブ定義でファイルシステムタイプリポジトリコネクションを選べると標準のものに加えて次の2つのタブが現れます。「ホップフィルタ」タブと「パス」タブです。

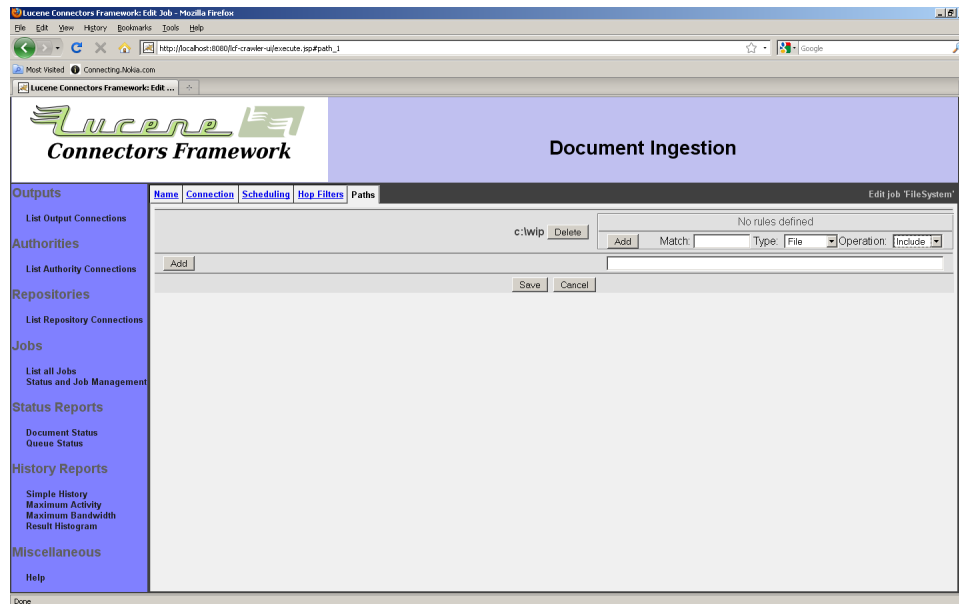
「ホップフィルタ」タブからは、ファイルを取得するサブディレクトリの深さ指定することができます。ファイルシステムの場合は、使われる場合は少ないかも知れませんが、この設定はWebコネクションタイプでも指定できます。ファイルシステムでこの設定の動作を確認することができます。



ファイルシステムコネクションタイプの場合は、コンテンツ間の関係は「子」の一つに限られています。コンテンツを取得するサブディレクトリのルートディレクトリからの深さを指定することが出来ます。空白の場合は、フィルタは無効と見なされます。

同じページから、コンテンツが存在するサブディレクトリの深さが変わった場合の処理を指定することができます。「読込めないコンテンツを削除」を選択すると、変更が発見されると、関係する可能性がすべてのコンテンツの深さを再計算します。再計算するとサーバに負荷が掛かります。再計算を行わないようにする場合は、一時的に行わないようにする設定と、永久に行わない設定があります。永久に行わないを選択すると、情報を削除します。

「パス」タブを選択すると、以下のようなページが表示されます：



このページからコンテンツを読込むパスを指定することができます。パスを入力して「追加」ボタンを押下するとパスが一覧に追加されます。パスの指定はManifoldCFが動作しているOSの形式で入力してください。

ルートパス毎に、コンテンツがジョブに含まれているかを判断するルールがあります。ルートパスを一覧に追加後にルールを定義することができます。ルール毎に、一致条件式、ファイルまたはディレクトリを対象にするかのフラグ、一致した場合にコンテンツを含むか除外するかを指定することができます。ルールは上から下に評価されます。最初にファイル名に一致したルールが適用されます。ルールを追加するには、プルダウンからタイプを選択して、一致する条件を入力してください(例:\*.txt)。入力後に「追加」ボタンを押下してください。

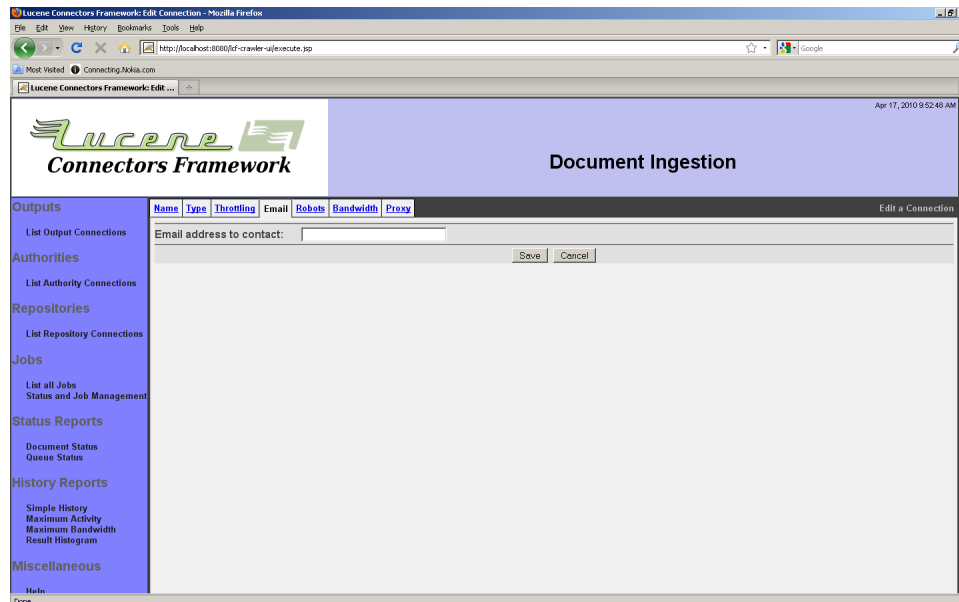
## 5.2 汎用RSSリポジトリコネクション

RSSコネクションタイプは、RSSフィードから索引を作成する場合に使います。Webコネクションタイプを利用してRSSフィードから索引を作成することもできますが、RSSコネクションタイプは以下の機能があります:

- フィードのみからリンクを抽出する
- フィード本体からは索引を作成しない
- フィードを再取得する条件を細かく指定することができる。また、通常のコンテンツとは異なる方法で処理される
- RSSコネクションタイプは特定のデータをメタデータとしてフィードからコンテンツに関連付ける

多くの場合、RSSコネクションタイプを利用するジョブは、継続的に実行され、コンテンツを再読み込みないように設定し、30日後にコンテンツを無効にします。この設定はニュースのRSSフィードから索引を作成する場合によく使われます。

RSSコネクションには4つの固有タグがあります:「メール」、「ロボット」、「バンド幅」、「プロキシ」。「メール」タブを選択すると以下のようなページが表示されます:

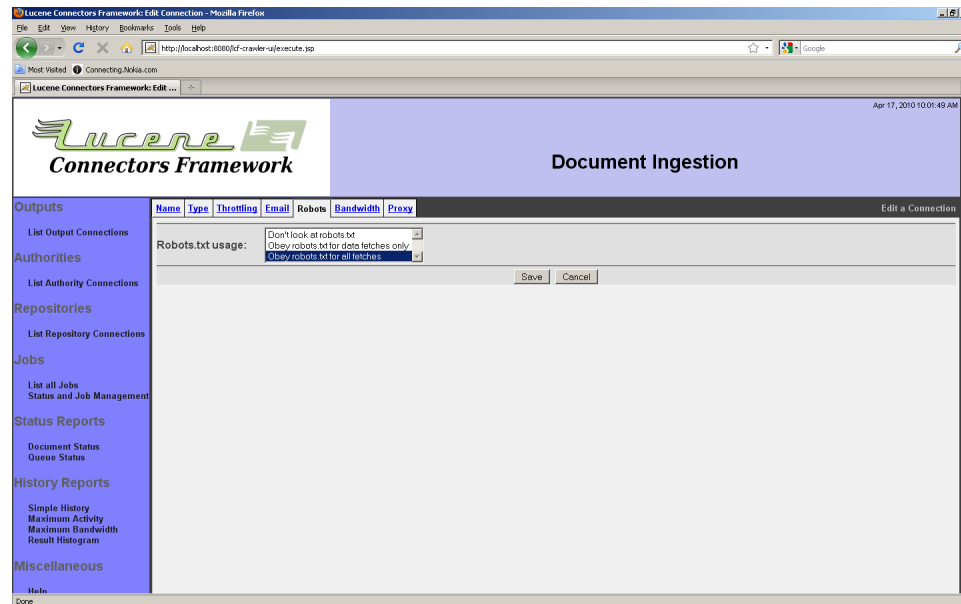


メールアドレスを入力して下さい。入力されたメールアドレスは、RSSコネクションのすべてのリクエストに含まれ、サーバ管理者が見ることができます。もし、スロットリング設定が大きすぎる場合でサーバ負荷が大きすぎる場合は、サーバ管理者からこのメールアドレスに連絡がされる可能性があります。

この項目は必須です。RSSコネクションはメールアドレスは妥当性を検証しませんが、ウェブ住民として正しいアドレスを入力してください。なお、サーバ管理者は「悪い」リクエストを拒否するように設定を変えることができますので、相手側サーバのことも考えてスロットリング設定を行ってください。

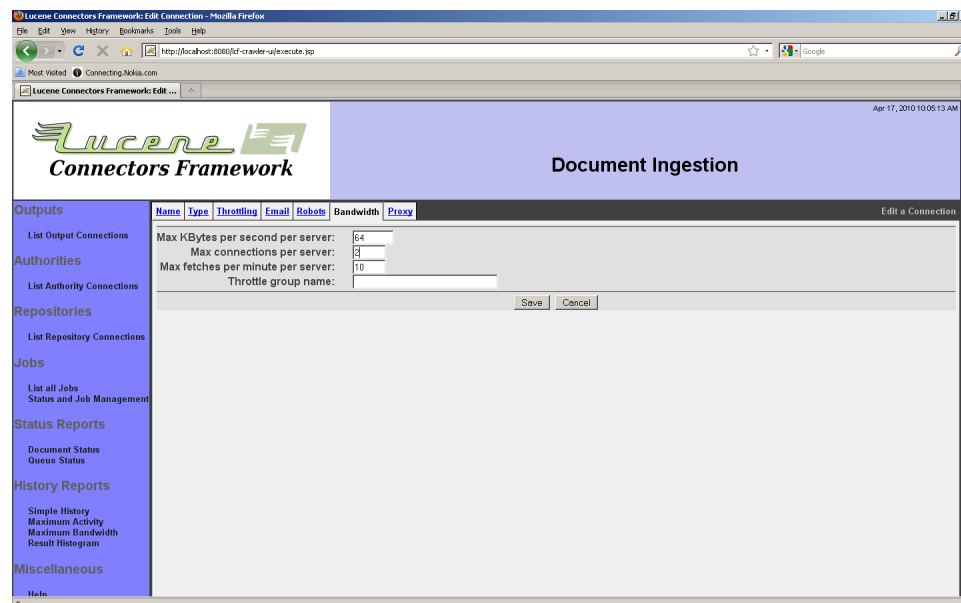
「ロボット」タブを選択すると次のようなページが表示されます:





robots.txtをどのように処理するかをプルダウンリストから選択してください。相手側サーバのことを考慮して選択してください。

「バンド幅」タブを選択すると次のようなページが表示されます：



サーバ毎に、接続がデータを取得する最大転送率及びサーバ毎に1分毎の最大転送率を設定できます。サーバ毎の最大ソケット接続数も指定できます。

ページの設定例の値は親切な設定になっています。デフォルトではすべての設定が空白ですので注意してください。このデフォルト設定では、スロットリングがされず、サーバに負荷を掛け、迷惑を掛けます。

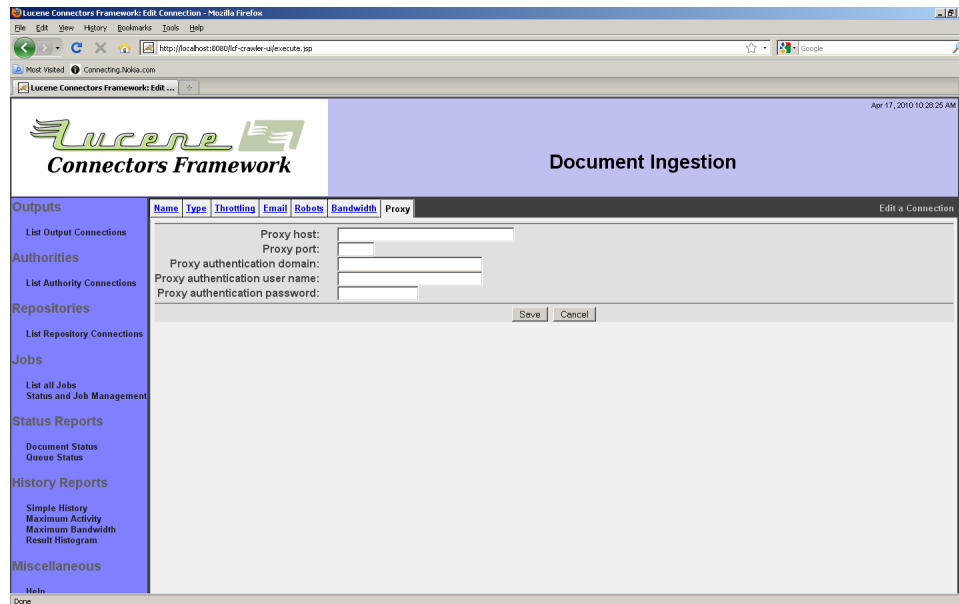
「スロットリンググループ」は、複数のRSSタイプコネクションのスロットリング設定を1つに纏めるための設定です。スロットリンググループ名が同じRSSタイプコネクションは同じスロットリングプールに纏められます。

「バンド幅」タブは「スロットリング」タブは次のように違います：

- 「バンド幅」タブからは最大値を設定できます。「スロットリング」タブからは平均値を設定することができます。
- 「バンド幅」タブからはコンテンツがどのようにキューにスケジュールされるかは設定できません。ただ、キューへのスケジューリングを遅らせるだけです。この待ち時間の間でもスレッドは使われます。「スロットリング」タブはコンテンツのジョブスケジューリングを行うため、待ち時間でスレッドを無駄に使いません。

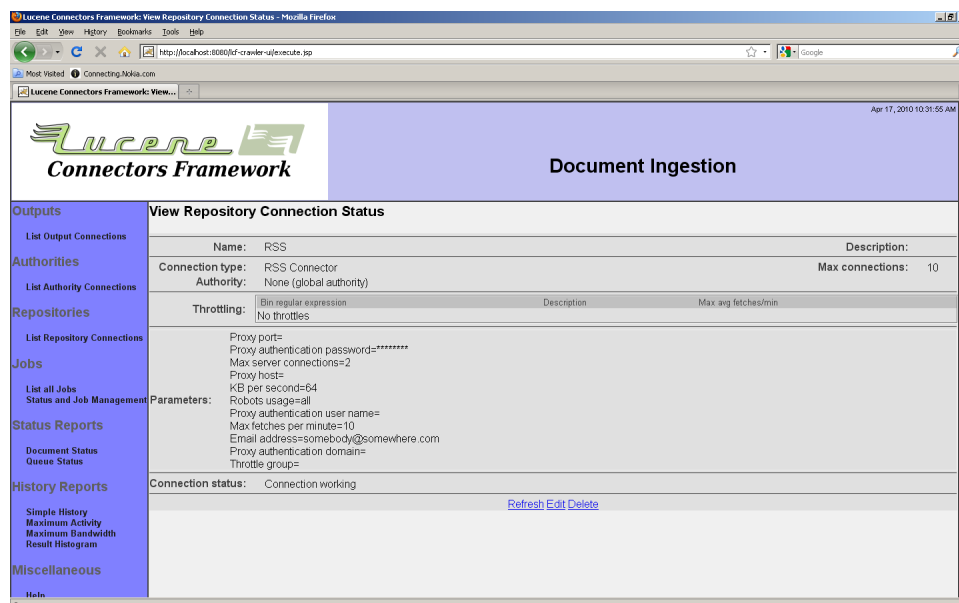
そのような理由のため、RSSコネクションには「バンド幅」タブと「スロットリング」タブの両方を設定することを推奨します。最大転送率を「バンド幅」タブで設定し、平均転送率を「スロットリング」タブで設定します。RSSコネクションのコンテンツIDはコンテンツのURLです。URLのbin名はサーバ名です。なお、「最大コネクション数/JVM」はデフォルトでは10です。この値は、RSSコネクションタイプには最適ではない可能性が高いです。ワークスレッド毎にコネクションを1つ設けることを推奨します。デフォルトではワークスレッド数は30ですので、「最大コネクション数/JVM」を30に設定にすることを推奨します。

プロキシを利用する場合は、「プロキシ」タブからプロキシ情報を入力してください。RSSコネクションタイプはNTLM認証のプロキシに対応しています。「プロキシ」タブを選択すると次のようなページが表示されます：

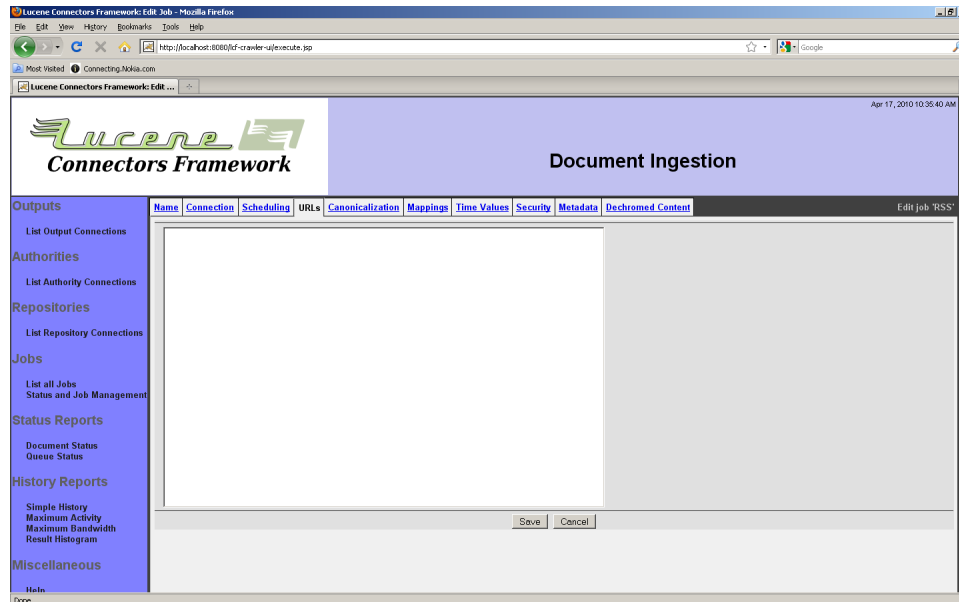


「プロキシホスト」にはプロキシサーバアドレスを入力してください。「プロキシポート」にはプロキシのポート番号を入力してください。認証が必要な場合は、ドメイン名、ユーザ名、パスワードを入力してください。プロキシを利用しない場合は、プロキシ関連のすべての項目を空にして置いてください。

RSSコネクションを保存すると、以下のような状態が表示されます。



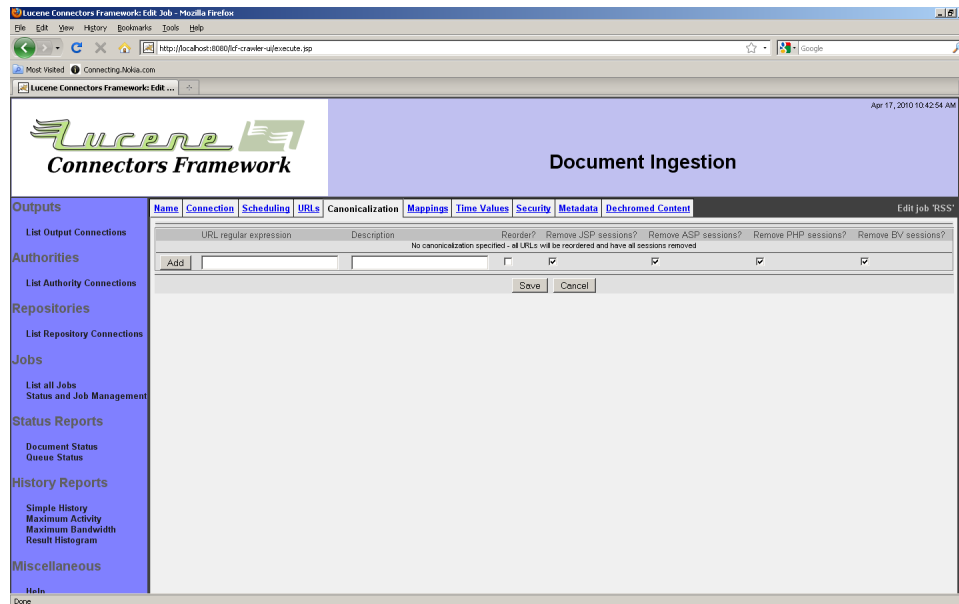
ジョブでRSSコネクションタイプを選択すると、次のタブが表示されます:「URL」、「正規化」、「マッピング」、「時間」、「セキュリティ」、「メタデータ」、「索引対象」。「URL」タブからジョブで対象にするRSSフィードの情報を指定します。「URL」タブを選択すると次のようなページが表示されます:



読込むRSSフィードのURLを改行で区切って入力してください。コメントを記入する場合は、行の先頭に「#」文字を入れて下さい。

「正規化」タブからはジョブがどのようにURLの正規化を処理するかを指定することができます。同一コンテンツに異なるURIが付けられている場合もあります。「正規化」機能は、このようなURLを同じURIと見なすために使います。例えば、URIの引数の順が異なっても同じコンテンツを指します: a=1&b=2とb=2&a=1は同じコンテンツを指すはずです。その他にもURIにセッションクッキー情報の有無もあります。

「正規化」タブを選択すると、次のようなページが表示されます:

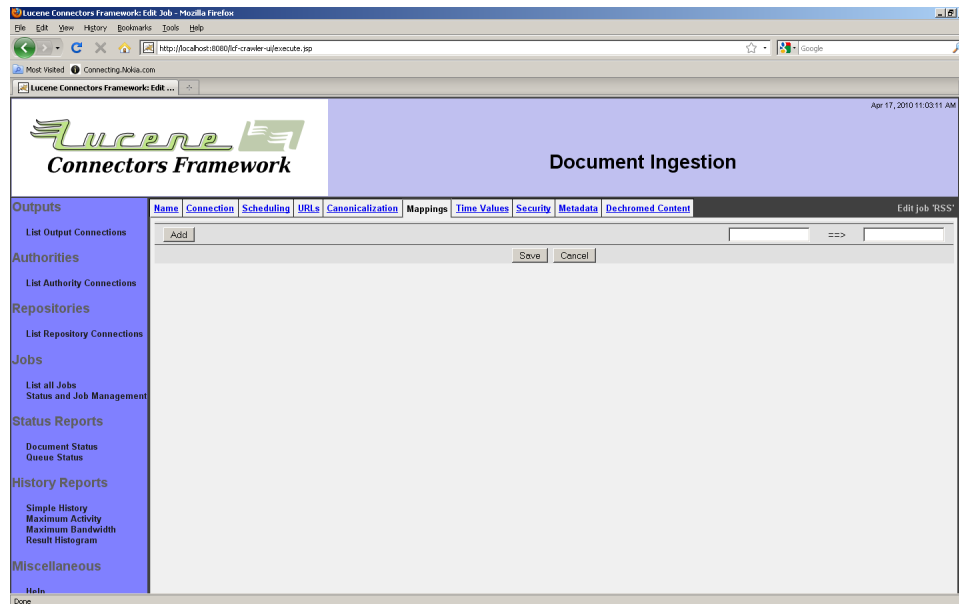


正規化するルール一覧が表示されます。各ルールは正規表現(対象URIを検索する)と条件項目から構成されます。条件項目で引数順の有無やセッションクッキー情報の排除などを指定することができます。次のセッションクッキー情報を排除することが出来ます: JSP (Javaアプリケーションサーバ)、ASP (.NET)、PHP、Broadvision (BV)。

ルールが複数のルールに一致する場合は、最初に一致したルールが適用されます。

ルールを追加するには、正規表現を入力して、条件項目のチェックボックスをチェックした後に、「追加」ボタンを押下してください。

「マッピング」タブから取得するコンテンツのURIを変更することが出来ます。例えばイントラネットのコンテンツを取得する場合に、一般ユーザが利用するURIと異なるURIを使ってコンテンツをクロールすることもあります。「マッピング」タブを選択すると次のようなページが表示されます:



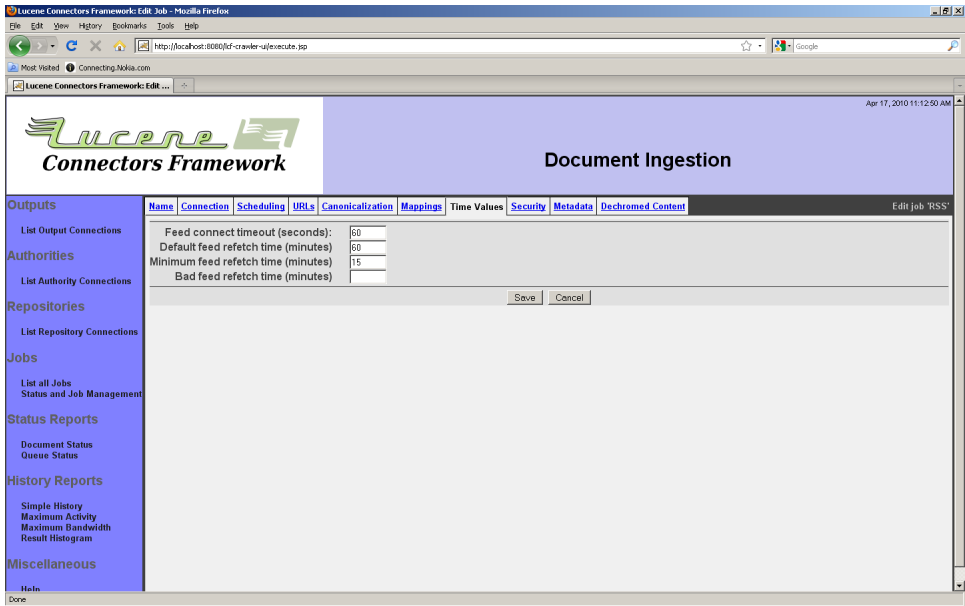
「マッピング」タブからはManifoldCFの他のタブで使われているのと同じ正規表現の仕組みで文字列の置き換えを設定することができます。マップはルールから構成されます。各ルールは一致する正規表現の式で構成されます。変換元と値は括弧(「(」と「)」)で囲みます。括弧に囲まれた部分を「グループ」と言います。置き換え文字列は、固定文字と置き換えグループから構成されます。例えば、「\$(1)」は最初に一致したグループを示し、「\$(1l)」は最初に一致した小文字のグループを示します。同じように「\$(1u)」は大文字にマップしたグループを示します。

例えば、ルール`http://(.*)/(.*)/`と置き換え文字列`http://$(2)/`は、`http://Server/Folder_1/Filename`を`http://Folder_1/Filename`に置き換えます。

1つ以上のルールが存在する場合は、上から実行され、上のルールの結果は下のルールで変更されます。

ルールを追加するには、一致する条件と出力文字列を入力して「追加」ボタンを押下してください。

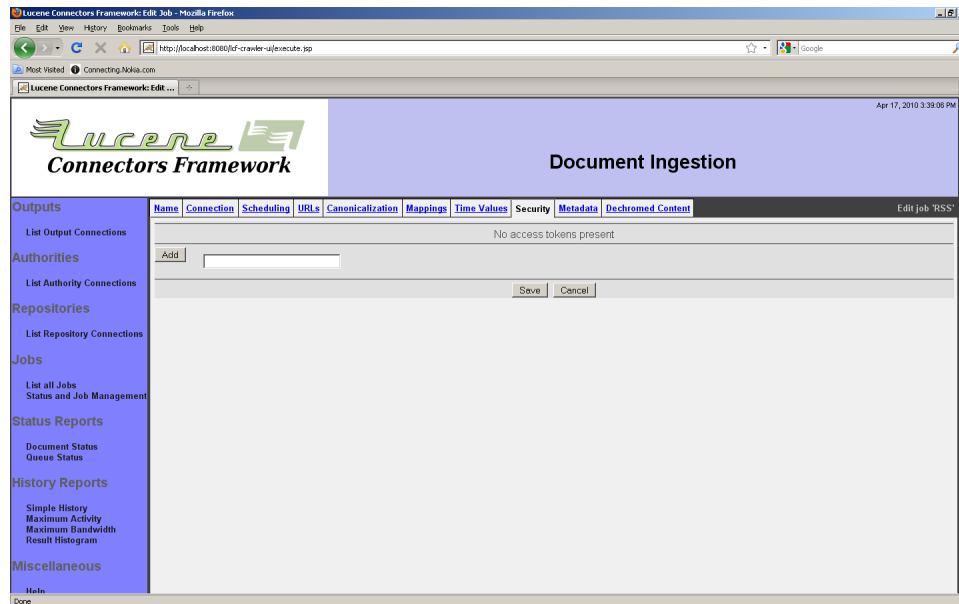
「時間」タブを選択すると次のようなページが表示されます：



設定したい時間制限値を入力してください。以下は項目の説明です：

値	説明
フィードタイムアウト	サーバ接続に待つ時間 (秒)
デフォルトフィード再取得時間	フィードに再取得時間が設定されていない場合に使う時間 (分)
最低再取得時間	フィードに設定した時間とは関係なく、設定した時間よりも短時間でフィードを再取得しない時間 (分)
エラーフィード再取得時間	解析エラーになったフィードを再取得するまでの待ち時間 (分、空の場合は無限)

「セキュリティ」タブからは、ジョブが利用する認証情報を設定することができます。利用する前に、利用する権限コネクションを決める必要があります。「セキュリティ」タブを選択すると次のようなページが表示されます：



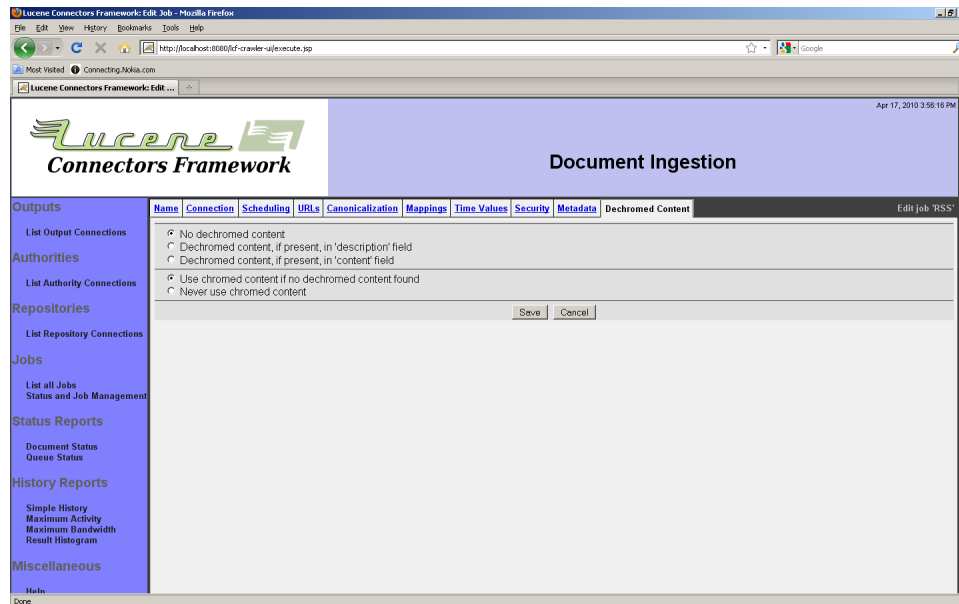
アクセストークンを追加するには、アクセストークンの値を入力して「追加」ボタンを押下してください。アクセストークンが無い場合は、ジョブのセキュリティは無効とされます。

「メタデータ」タブからは、ジョブのすべてのコンテンツからの索引に添付するメタデータを指定することができます。RSSコネクションタイプのコンテンツは、以下のような標準でメタデータが付けられます：

名前	説明
PubDate	コンテンツが作成された日時（1970年1月1日からのミリ秒）。作成日を取得できない場合は、コンテンツを取得した日時になります。
Source	コンテンツの名前。
Title	フィード内のコンテンツの題名。
Category	フィード内のコンテンツの分類。

「索引対象」タブからは、コンテンツからではなく、フィードの概要から索引を作成するように指定することができます。内容がリンク一覧のようなフィードで、フィードの概要から索引を作成する場合に使うことができます。「索引対象」タブを選択すると次のようなページが表示されます：





コネクションで利用するモードを選択してください。

### 5.3 汎用Webリポジトリコネクション

Webコネクションタイプは、Webクローラです。基本認証、NTLM認証、セッション認証に対応しています。以下のようなコンテンツを処理することができます：

- テキスト
- HTML
- 汎用XML
- RSSフィード

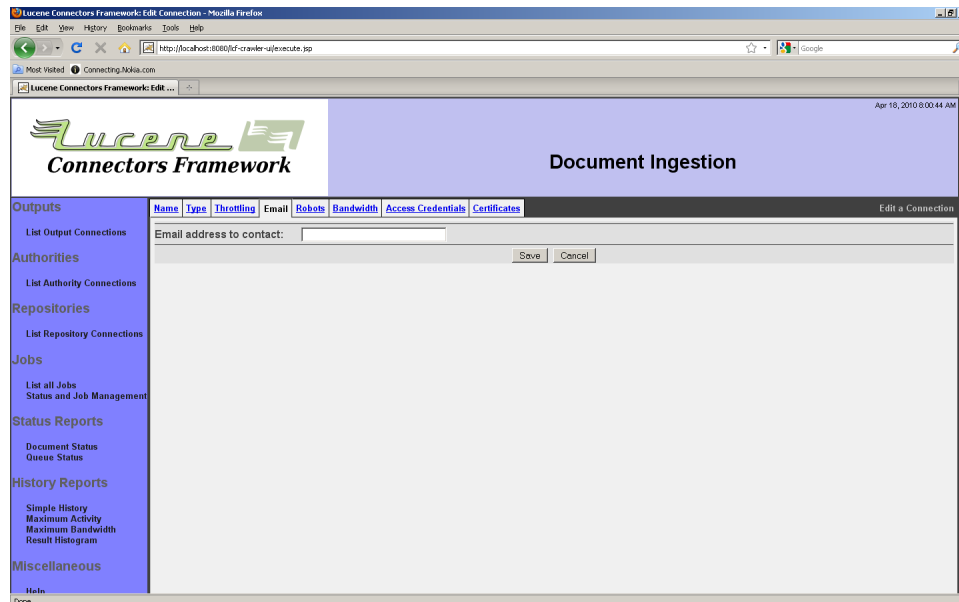
WebコネクションタイプはRSSコネクションタイプと以下の機能が異なります：

- 出力コネクションがフィードを受け付けた場合は、フィードから索引は作成されます。
- すべてのコンテンツからリンクも抽出されます。
- フィードは他コンテンツと同じように処理されます。一つだけの再取得設定を行うことはできません。
- ホップ数による制限を設けることができます。
- URIセットに含める／除外することができます。

WebコネクションタイプはRSSコネクションタイプよりも設定が複雑で、RSSフィードの詳細設定を行うことはできません。その結果、RSSの索引を作成する場合は、RSSコネクションタイプを利用することを推奨します。

Webコネクションタイプを利用する多くのジョブは、継続的に実行され、定期的にコンテンツを再取得するか、コンテンツを一回限り取得して再取得しないように設定され、指定した期間後に無効になるように設定されます。

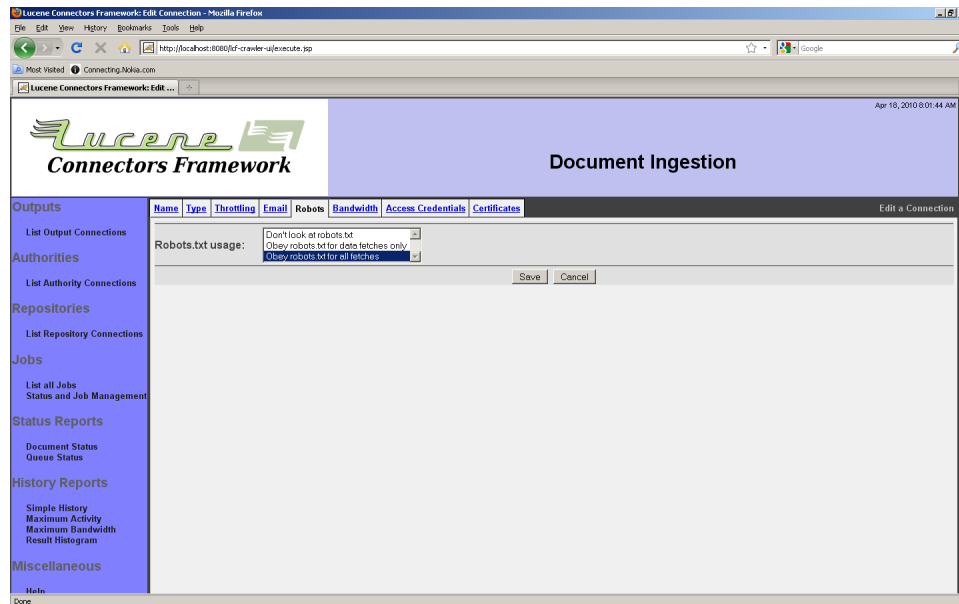
Webコネクションタイプを選択すると次のタブが表示します:「メール」、「ロボット」、「バンド幅」、「認証」、「証明書」。「メール」タブを選択すると次のようなページが表示されます:



メールアドレスを入力して下さい。入力されたメールアドレスは、Webコネクションのすべてのリクエストに含まれ、サーバ管理者が見ることができます。もし、スロットリング設定が大きすぎる場合でサーバ負荷が大きすぎる場合は、サーバ管理者からこのメールアドレスに連絡がされる可能性があります。

この項目は必須です。Webコネクションはメールアドレスは妥当性を検証しませんが、ウェブ住民として正しいアドレスを入力してください。なお、サーバ管理者は「悪い」リクエストを拒否するように設定を変えることができますので、相手側サーバのことも考えてスロットリング設定を行ってください

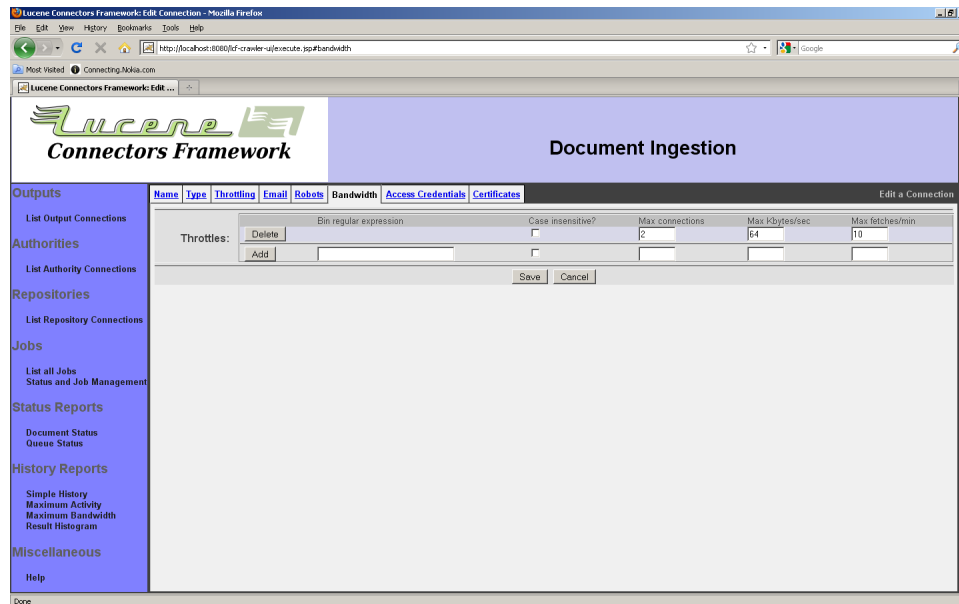
「ロボット」タブを選択すると次のようなページが表示されます:



robots.txtをどのように処理するかをプルダウンリストから選択してください。相手側サーバのことを考慮して選択してください。

「バンド幅」タブからはバンド幅ルール一覧を設定することができます。ルール毎にURLスロットルbinを選択する正規表現を指定します。WebタイプのスロットルbinはURIのサーバ名です。ルール毎に最大バンド幅、コネクション数、読み込み率を指定することができます。任意の数だけルールを作成することができます。もしURLが複数のルールと一致した場合は、一番保守的なルールが利用されます。

「バンド幅」タブを選択すると、次のようなページが表示されます：



ページの設定例の値は親切な設定になっています。デフォルトではすべての設定が空白ですので注意してください。このデフォルト設定では、スロットリングがされず、サーバに負荷を掛け、迷惑を掛けます。

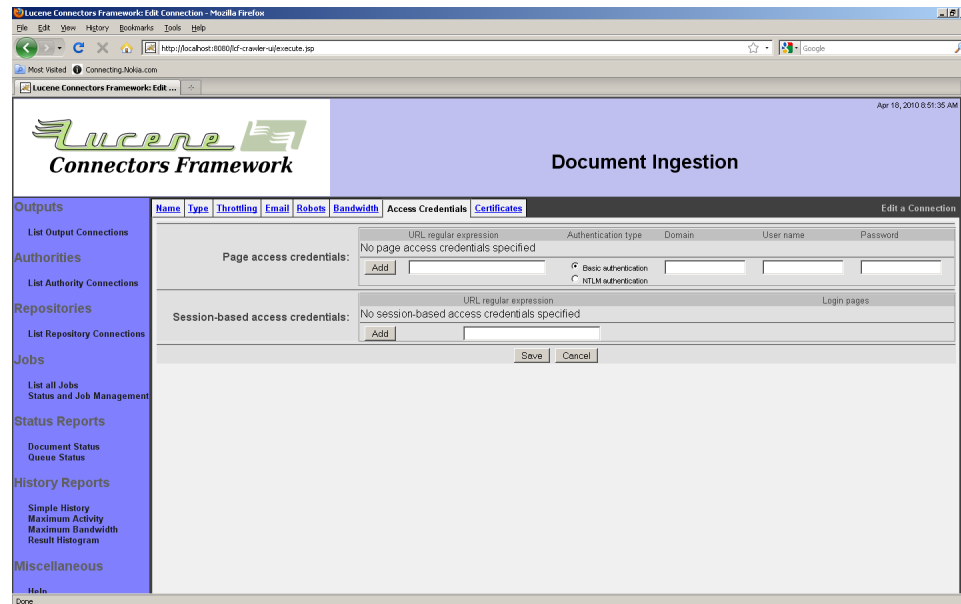
ルールを追加するには、正規表現と制限値を入力して「追加」ボタンを押下してください。

「バンド幅」タブと「スロットリング」タブは次のように違います：

- ・「バンド幅」タブからは最大値を設定できます。「スロットリング」タブからは平均値を設定することができます。
- ・「バンド幅」タブからはコンテンツがどのようにキューにスケジュールされるかは設定できません。ただ、キューへのスケジュールを遅らせるだけです。この待ち時間の間でもスレッドは使われます。「スロットリング」タブはコンテンツのジョブスケジュールリングを行うため、待ち時間でスレッドを無駄に使いません。

そのような理由のため、Webコネクションには「バンド幅」タブと「スロットリング」タブの両方を設定することを推奨します。最大転送率を「バンド幅」タブで設定し、平均転送率を「スロットリング」タブで設定します。WebコネクションのコンテンツIDはコンテンツのURLです。URLのbin名はサーバ名です。なお、「最大コネクション数/JVM」はデフォルトでは10です。この値は、Webコネクションタイプには最適ではない可能性が高いです。ワーカスレッド毎にコネクションを1つ設けることを推奨します。デフォルトではワーカスレッド数は30ですので、「最大コネクション数/JVM」を30に設定にすることを推奨します。

Webコネクションの「認証」タブからはページ認証方法を指定することができます。ページ認証（例：基本認証、NTLM認証）及びセッション認証（ログインセッション）に対応しています。「認証」タブの初期ページには両方の認証方法が表示されています：



認証方法毎にルール一覧を設けることができます。

ページ認証を設定する場合は、認証を必要なURI、認証方法とそのユーザ／パスワードを指定します。これらの項目を入力した後に「追加」ボタンを押下してください。

セッション認証を設定する場合は、少し調べる必要があります。セッションで保護されているサイト毎にセッション認証ルールを設ける必要があるかもしれません。サイト毎に次のような情報が必要です：

- セッションセキュリティで保護されているページのURI。
- ログイン処理中のページの取得する手順。
- ログインページにログイン情報の入力方法。

Webコネクションはログイン処理中のページを「login pages」とし、保護されているページを「content pages」とします。Webコネクションはログインページの索引を作成しません。ログインページは認証情報の入力用のページでコンテンツ情報が含まれていないからです。

また、サイトに初めて訪問する場合と、セッションが無効になってログインしていなくてコンテンツを取得しようとする場合も考慮する必要があります。両方の場合は、セッション認証ルールを適用してコンテンツを取得する必要があります。ManifoldCFフレームワークでは何時コンテンツを取得または再取得するか制御はできません。

ログインページのURI及び特徴的な内容からログインページを示します。例えば、セッションが無効になった場合はログインページにリダイレクトするサイトもあります。このような場合は、コンテンツを取得するよりも、ログインページへのリダイレクト情報を取得します。一般的には、ログインページ及びリダイレクト情報をコンテンツと区別して索引を作らないようにします。このような場合は、3つのログイン情報を登録します：一つはログインページへのリダイ

レクト、もう一つはログインページのURL、最後の一つはログインフォームの送信先。ログインページにログイン情報を設定して、送信するようにします。

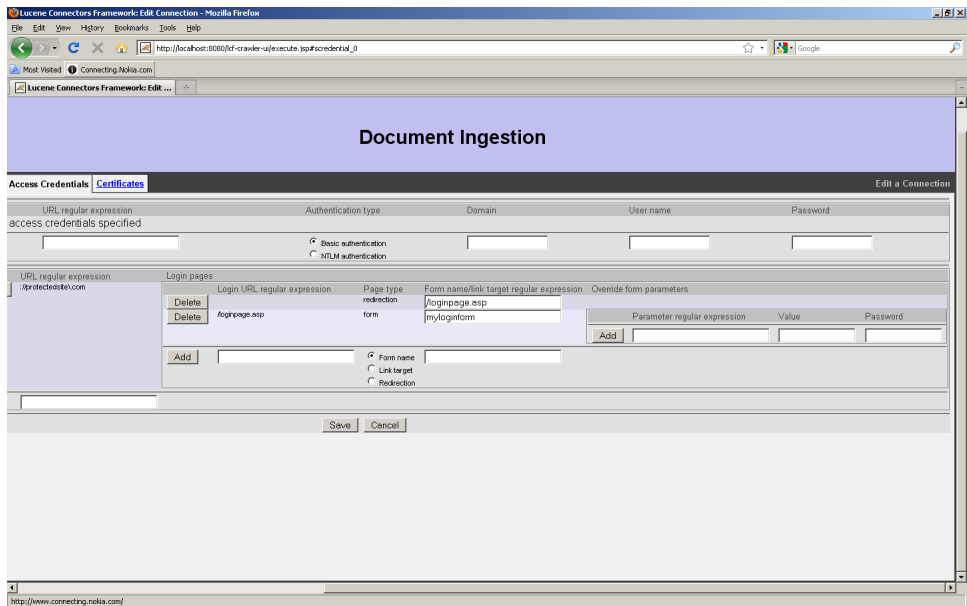
Webコネクションは次のような内容をログインページと見なすことができます：

- 特定のURI(正規表現に一致)へのリダイレクト
- 指定した名前(正規表現に一致)のフォーム(FORM)が存在するページ
- 特定のページへのリンク(正規表現に一致)を含んだページ

セッション認証ルールを追加する場合は、保護されたページを特定する正規表現を入力して「追加」ボタンを押下します。次のようなページが表示されます：

新規に作成されたルールにログインページの情報を入力することができます。ログインページ情報を入力するには、URI正規表現、ログインページタイプ、ターゲットリンクまたはフォーム名正規表現を入力して、「追加」ボタンを押下してください。

「フォーム」型のログインページを追加した場合は、次のようにフォームにログイン情報を入力することができます：



フォームの項目に入力する値を設定してください。入力内容を非表示にする場合は、「値」列の代わりに「パスワード」列に値を入力してください。フォームの項目名はログインページのHTMLソースコードを表示して、調べてください。入力した後に「追加」ボタンを押下してください。

指定されていないフォーム項目はログインページのデフォルト値で送信されます。現バージョンでは、Javascriptは未対応です。ログインフォームにJavascriptが利用されている場合は、スクリプトの結果を事前に計算して、その結果を登録してください。複雑なJavascriptを含むログインページの場合は、設定値を探すのに時間が掛かる場合があります。

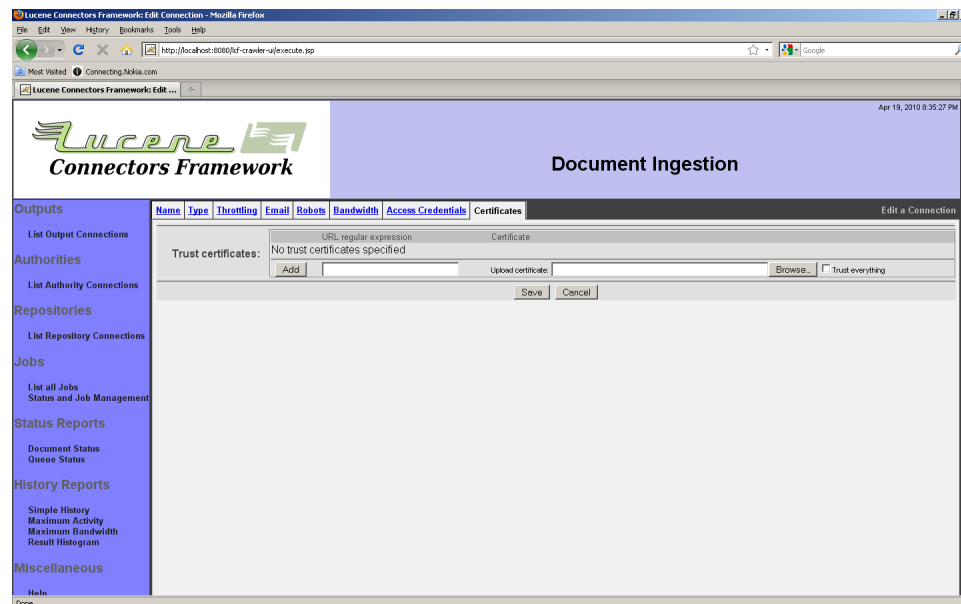
複数のログインページフォームはサイトの「ログインページ手順」です。Webコネクションは、ログインページ毎に次に取得する内容をログインページ条件で決めます。例えば、特定のURIへのリダイレクトの場合は、リダイレクトURIを取得します。フォームの場合は、フォームのactionで指定したページを取得します。ターゲットへのリンクの場合は、ターゲットURIを取得します。ログインページ手順の最後にはWebコネクションがログイン手順を開始する前に元々取得するページを取得します。

セッション認証をデバッグする場合は、Webコネクションの履歴レポートを参照することを推奨します。Webコネクションのイベント履歴を参照することで大体の動作が分かるはずですが、以下のようなイベントがあります：

イベントタイプ	説明
Fetch	URIの取得履歴です。HTTPからの戻り値はレスポンスコードとして記録されます。HTTP処理が失敗または不完了のイベントは負の値で記録されます。

Begin login	ログイン手順を実行する場合に記録されます。ログイン手順が実行されると、ログインが完了するまでは他の保護されたサイトからコンテンツは取得されません。
End login	ログイン手順から元のコンテンツの取得に戻った時に記録されます。元のコンテンツの取得に戻った場合は、サイトから並行してコンテンツの取得を再開します。

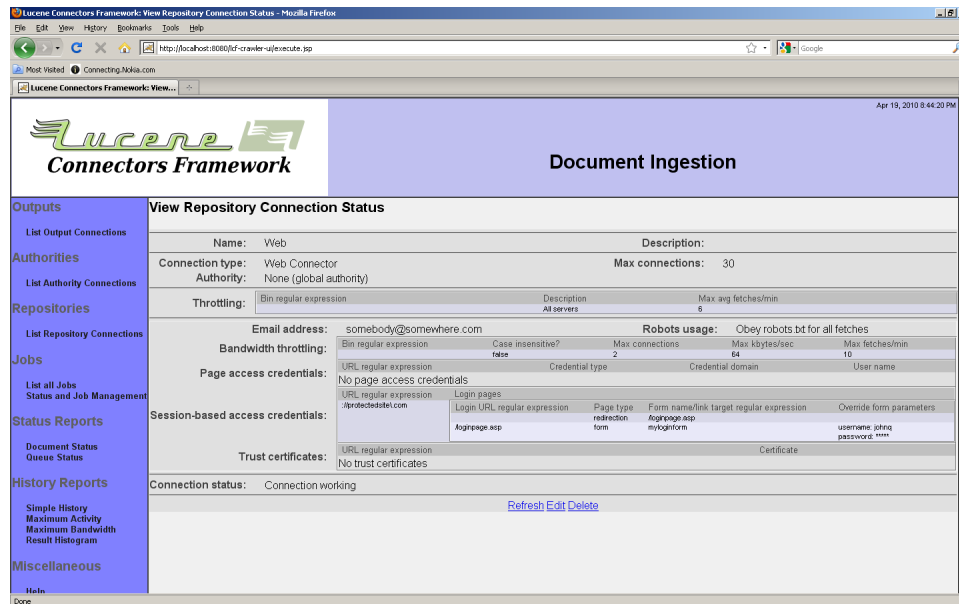
「証明書」タブはSSLと一緒に利用され、信用した正規表現と一致したURIの証明書情報を設定します。すべての証明書を信用することもできます。「証明書」タブを選択すると次のようなページが表示されます：



URI正規表現を入力し、「すべての信用する」チェックボックスをチェックするか、証明書を参照してください。（サーバの証明書を信用することもできますが、証明書が期限切れになる場合もあります。）証明書を一覧に追加する場合は「追加」ボタンを押下してください。

入力した後に「保存」ボタンを押下すると次のような設定内容の概要ページが表示されます：



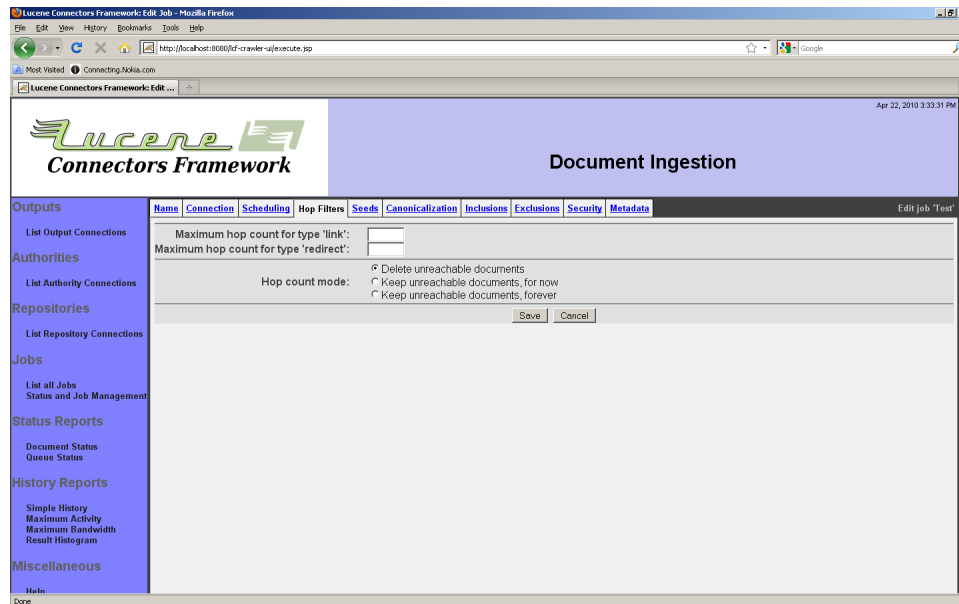


ジョブでWebタイプのリポジトリコネクションを選択した場合は、次のタブが表示されます：「ホップフィルタ」、「シード」、「正規化」、「含む」、「除外」、「セキュリティ」、「メタデータ」。

「ホップフィルタ」タブからは、シードコンテンツからの最大ホップ数を指定することができます。Webタイプのコネクションには2種類のホップ数があります：「リンク」ホップと「リダイレクション」ホップ。ホップの種類毎に最大数を設定することができます。空白の場合は、無限と見なされます。

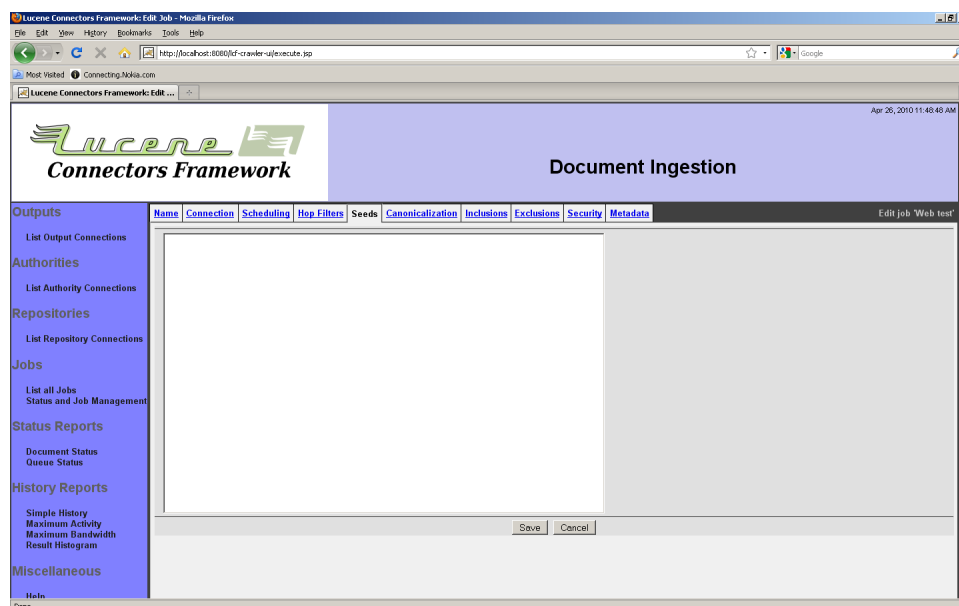
例えば、最大「リンク」ホップ数を5に設定して、「リダイレクト」ホップ数を空白にした場合は、シードコンテンツから5ホップより多いコンテンツは対象外とされます。最大「リンク」ホップ数を5に設定して、最大「リダイレクト」ホップ数を2に設定した場合は、シードコンテンツから5ホップより多くてかつリダイレクトのホップ数が2より多いコンテンツは対象外とされます。

「ホップフィルタ」タブを選択すると次のようなページが表示されます：

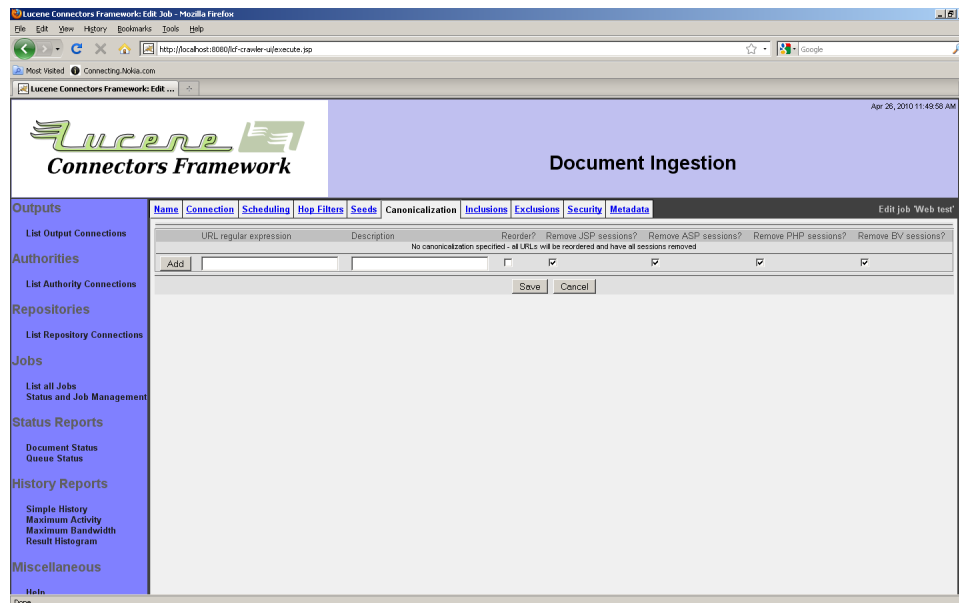


このページからルートからコンテンツのホップ数に変更された場合に行う処理を指定することができます。「読込めないコンテンツを削除」を選択すると、変更が発見されると、関係する可能性があるすべてのコンテンツの深さを再計算します。再計算するとサーバに負荷が掛かります。再計算を行わないようにする場合は、一時的に行わないようにする設定と、永久に行わない設定があります。永久に行わないを選択すると、情報を削除します。

「シード」タブからクロールを始めるコンテンツを指定します。「シード」タブを選択すると次のようなページが表示されます：



シードを改行で区切って入力してください。空行及び「#」から始まる行は無視されます。  
 「正規化」タブからURIを標準形式に変換するルールを入力することができます。「正規化」タブを選択すると次のようなページが表示されます：

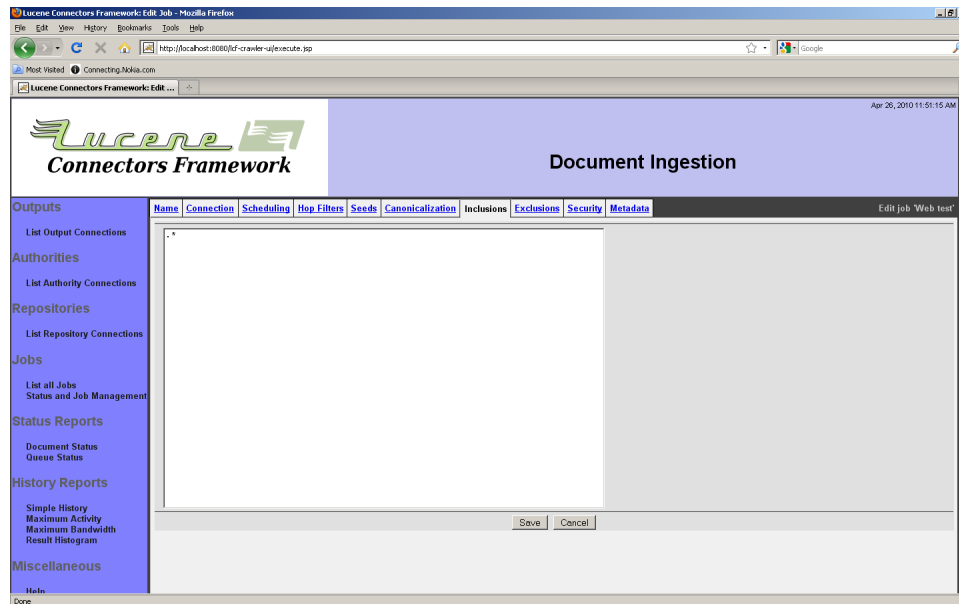


正規化するルール一覧を表示します。各ルールは正規表現(対象URIを検索する)と条件項目から構成されます。条件項目で引数順の有無やセッションクッキー情報の排除などを指定することができます。次のセッションクッキー情報を排除することが出来ます：JSP (Java アプリケーションサーバ)、ASP (.NET)、PHP、Broadvision (BV)。

ルールが複数のルールに一致する場合は、最初に一致したルールが適用されます。

ルールを追加するには、正規表現を入力して、条件項目のチェックボックスをチェックした後、「追加」ボタンを押下してください。

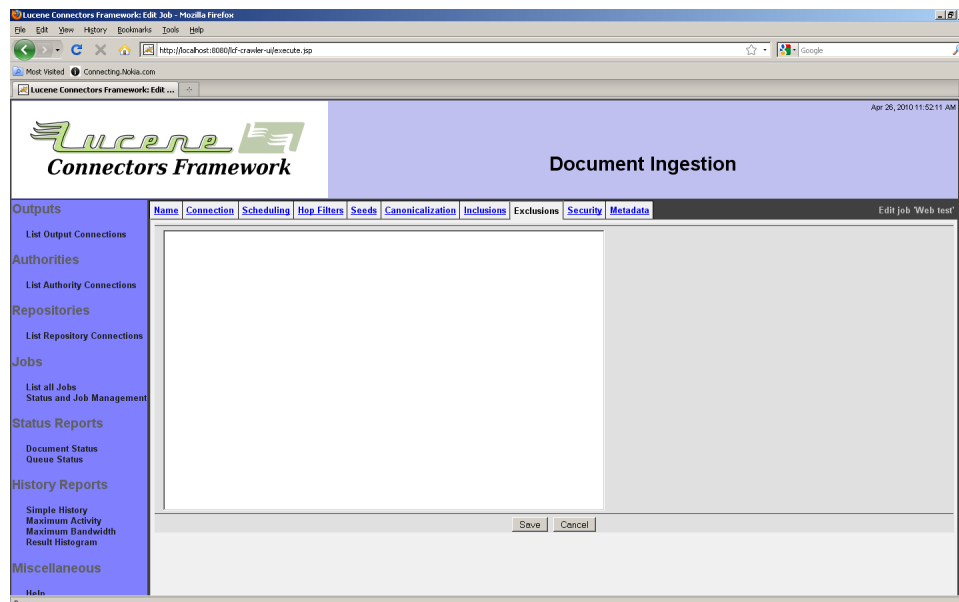
「含む」タブからWebジョブに含むURI正規表現一覧を指定することができます。「含む」タブを選択すると次のようなページが表示されます：



改行区切りで0以上の正規表現を指定してください。

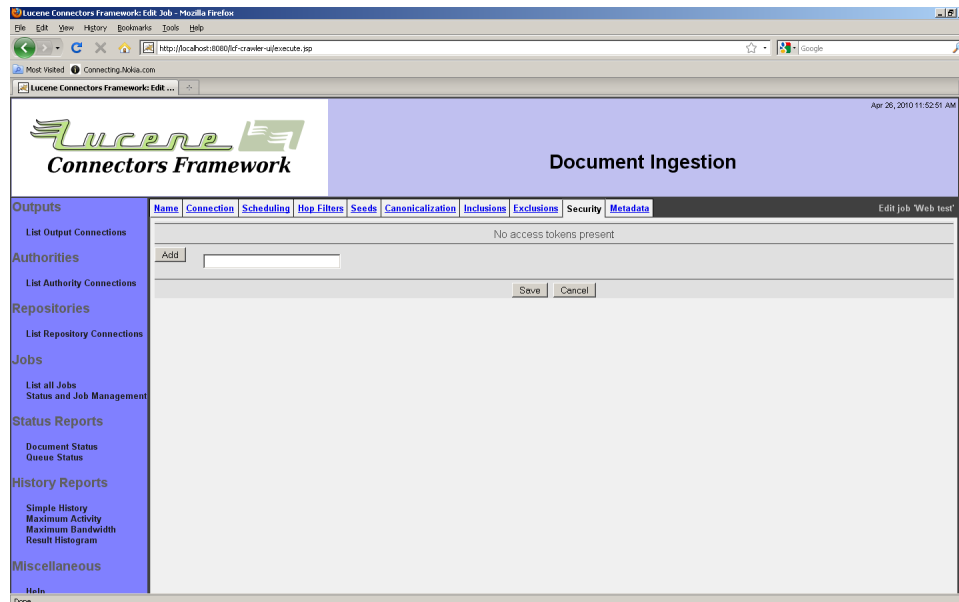
Webジョブはデフォルトでは、シードにリンクされているインターネット上のすべてのコンテンツを含みます。

対象にするコンテンツを制限したい場合は、「除外」タブから指定してください。「除外」タブを選択すると次のようなページが表示されます：



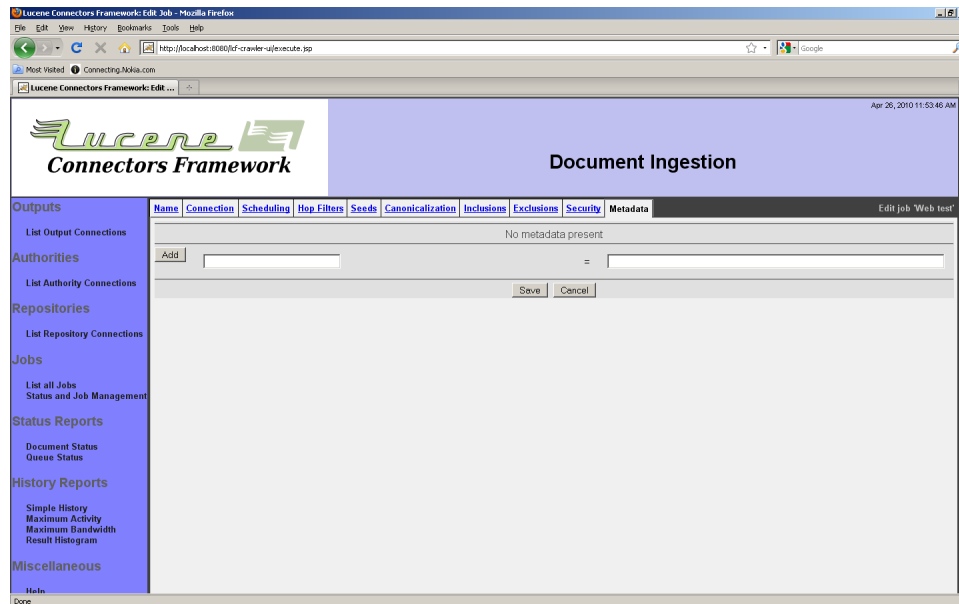
改行区切りで0以上の正規表現を指定してください。索引を作成できないコンテンツは、索引をする必要がないコンテンツを除外することを推奨します。例えば動画や映像などからは索引を作成できないので除外の対象です。

「セキュリティ」タブからWebジョブが索引を作成するコンテンツのアクセストークンを指定することができます。「セキュリティ」タブを選択すると次のようなページが表示されます：



コンテンツにセキュリティを追加する前に、アクセストークンの形式の情報が必要です。アクセストークンを入力して「追加」ボタンを押下してください。

「メタデータ」タブからコンテンツにメタデータを付けることができます。「メタデータ」タブを選択すると次のようなページが表示されます：



設定するメタデータ名と値を入力して「追加」ボタンを押下してください。

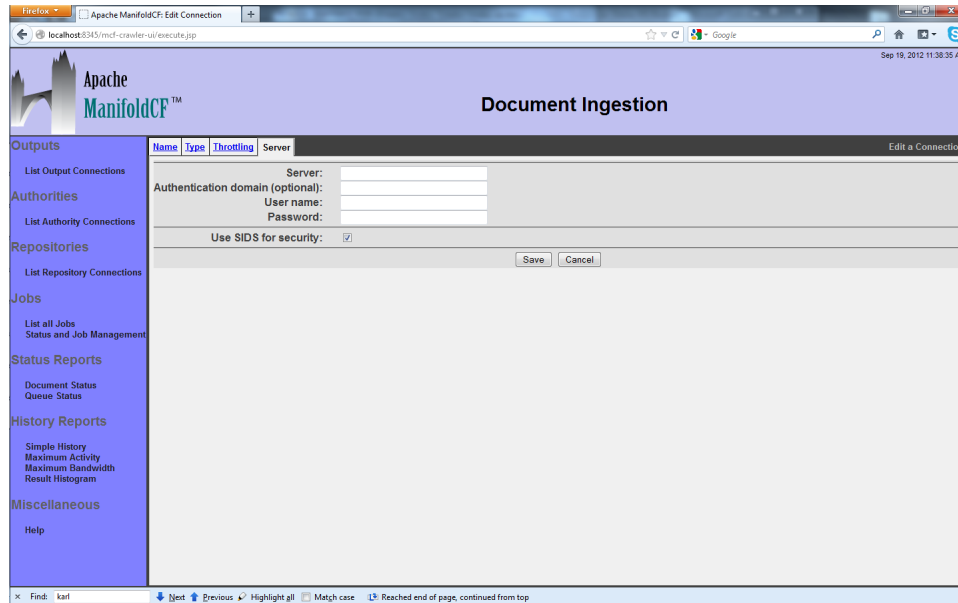
#### 5.4 Windows Share/DFSリポジトリコネクション

Windows共有コネクションタイプは、Windowsの共有フォルダにあるコンテンツを索引する場合に使います。Windows以外のOSのシステムから利用することができます。Sambaや第三者のNASサーバにも対応しています。

DFSノードと参照はすべて対応しています。ただし、参照するサーバ名はManifoldCFがインストールされたサーバが利用しているDNSから参照できることを前提とします。Windows共有コネクションは、コンテンツ毎に索引IDを作成します。索引IDは「file:」IRIまたは「http:」URIです。柔軟なデプロイ環境が可能ですが、設定に少し時間が必要になります。特にファイルIRIを利用する場合は、システムの検索コンポーネントが正しく対応しているのか確認してください。Internet ExplorerのようなWebブラウザからWindowsファイルシステムのコンテンツを閲覧する場合は、¥¥servername¥sharename¥dir1¥filename.txtのようなアドレスをfile://///servername/sharename/dir1/filename.txtのようなIRIに変換します。簡単のようですが空白、「#」、英数以外の文字がファイル名に含まれている場合は複雑になります。Internet Explorerのバージョンによって異なる方法で処理するため、一つの方法でWindows共有ファイルパスをIRIに変換することはできません。代わりにコネクションは標準正規化アドレスを利用して、システムが索引結果をWebブラウザ及びクライアントに正しい方法で変換することを期待します。

権限付きでWindows共有リポジトリコネクションでコンテンツをクロールする場合は、事前にActive Directory権限コネクタを作成してください。

Windows共有コネクションはリポジトリコネクション編集ページで1つの固有タブがあります:「サーバ」タブ。「サーバ」タブを選択すると次のようなページが表示されます:

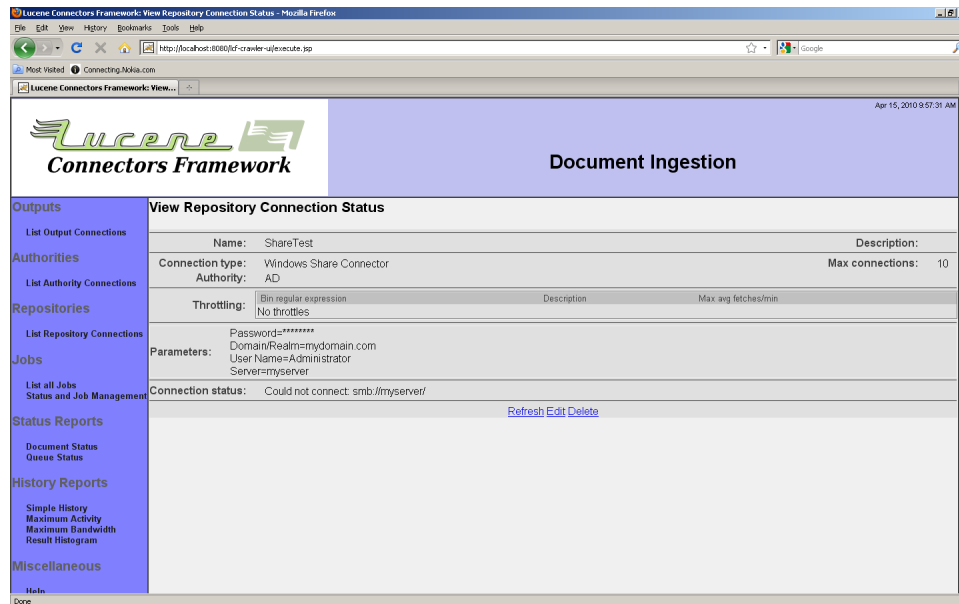


接続するサーバ名を入力してください。サーバ名は、実サーバ名またはWindowsドメインDFSルートに接続する場合はドメイン名で指定することができます。実サーバ名を入力する場合は、サーバ名に未修飾名を入力して、「ドメイン名」に完全修飾ドメイン名を入力してください。ユーザ名は未修飾名を入力して下さい(例:「Administrator@mydomain.com」ではなく、「Administrator」と入力してください)。「ドメイン名」を空白にして、「サーバ」に完全修飾サーバ名を入力する場合があります。ただし、ドメイン名と完全修飾サーバ名の両方を入力しないでください。

"Use SIDs"チェックボックスは、コネクションがSIDsをアクセストークンとして使うか(これはActive Directoryによるセキュリティ制御されたWindowsサーバやNASサーバに適しています)、あるいはユーザ/グループ名を使うか(これはSambaサーバや、LDAP権限コネクションタイプと連携してLDAPを使う他のCIFSサーバに適しています)を制御します。SIDsを使うならチェックしてください。

サーバ側の負荷を軽減するために、「スロットリング」タブの「最大コネクション数/JVM」をデフォルト値の10より少ない値に変更することを推奨します。Windowsはマルチスレッド処理をそんなに良くしません。5以下に設定してもそんなに取得性能は変わりません。

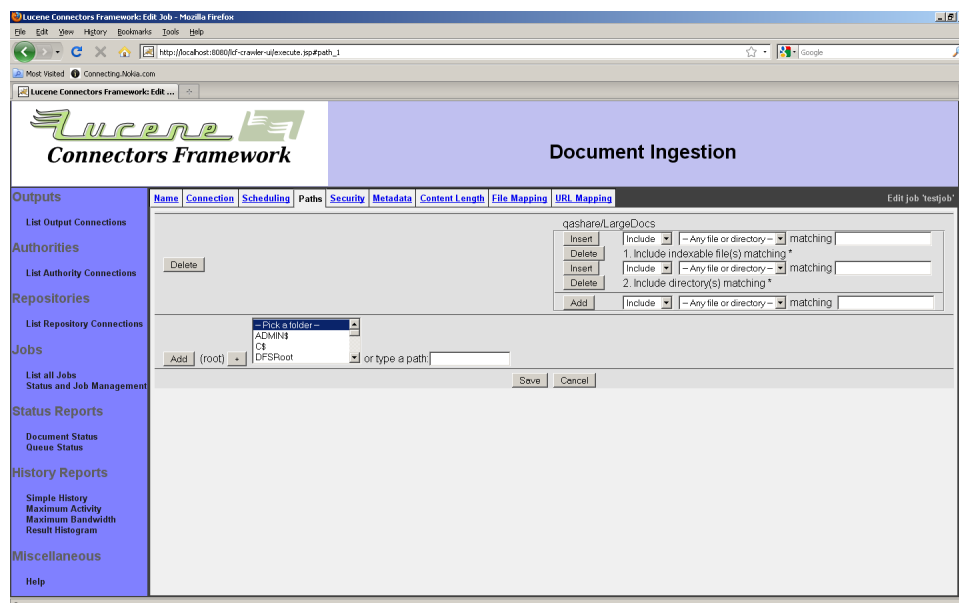
入力した後に「保存」ボタンを押下してください。次のような設定概要ページが表示されます:



画面例では、Windows共有コネクションはサーバに接続できないためエラーメッセージが表示されています。

ジョブでWindows共有タイプを選択すると次のタブが表示されます:「パス」、「セキュリティ」、「メタデータ」、「Content Length」、「ファイルマップ」、「URLマップ」。

「パス」タブを選択すると次のようなページが表示されます:





このページから起点パスの指定、生成パスの追加、既存パス一覧からパスの削除することができます。起点パスを指定しないと、ジョブが対象とするコンテンツはありません。

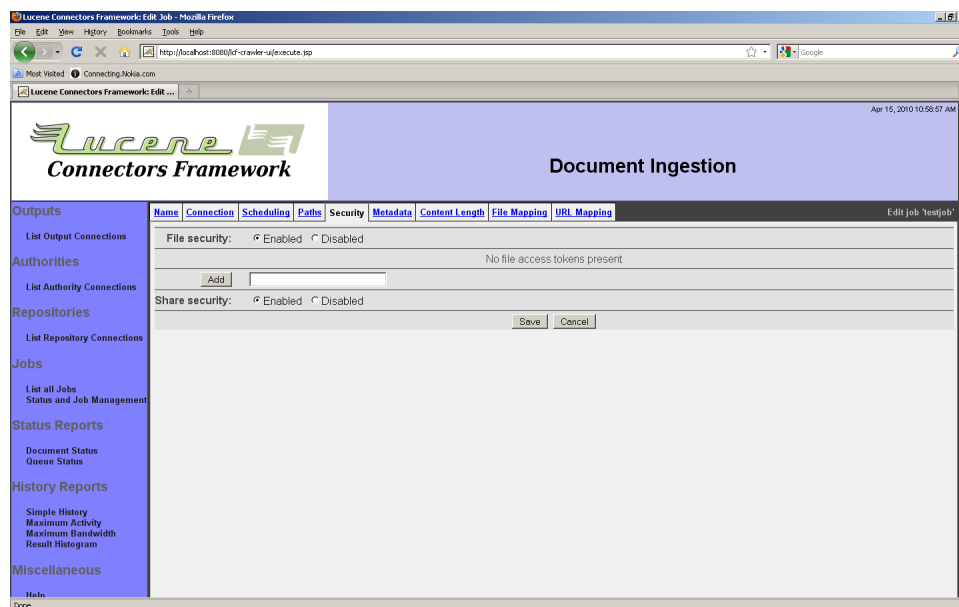
コネクションの状態が「起動」になっていることを確認してください。エラーメッセージが表示している場合は、パスは作成されません。

含むパス毎にジョブが対象とするフォルダ及びコンテンツを特定するルール一覧が表示します。ルールは上から下へ評価されます。最初に一致したルールが使われます。

各ルールはパスを特定する条件を指定します。各ルールはファイル名条件(例:「\*.txt」)、ファイルまたはフォルダ名を特定するのかの指定、出力コネクションでファイルから索引を作成するか否か、ファイルを含むか除外するかの指定が含まれます。ファイル名の指定にはワイルドカード文字「\*」と「?」を使うこともできます。「\*」は0以上の任意の文字と一致します。「?」は任意の1文字と一致します。その他の文字は記述通りに一致する必要があります。

起点パスのルールを追加する場合は、プルダウンメニューから値を選択して、ファイル選択条件を入力して、「追加」ボタンを押下してください。既にあるルールの上にルールを挿入する場合は「挿入」ボタンを押下してください。

「セキュリティ」タブを選択すると次のようなページが表示されます：



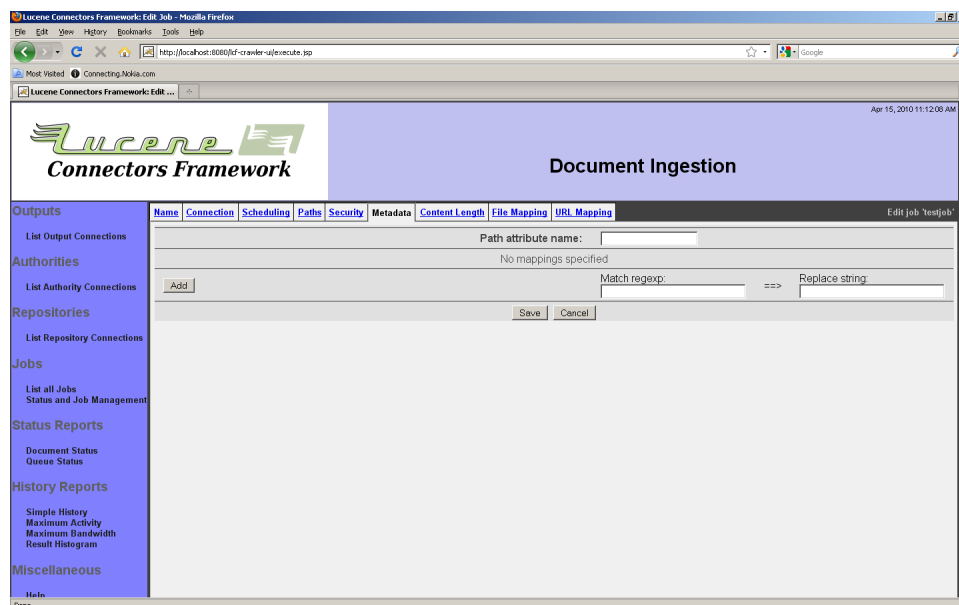
「セキュリティ」タブからは次の3機能を設定することができます：ファイルセキュリティ、共有セキュリティ、ジョブで索引が作成されたすべてのコンテンツのセキュリティトークン(セキュリティが無効の場合)。

ファイルセキュリティとは、Windowsが各ファイルに適用しているセキュリティです。ほとんどのWindows互換のNASタイプサーバでサポートされています。

共有セキュリティとは、WindowsがWindows共有フォルダで使っているセキュリティです。古いセキュリティの仕組みで、このセキュリティを利用しているユーザは少なくなっています。最新のNASシステムやSambaではサポートされていない場合があります。Windows共有セキュリティがサポートされていないシステムでこのコネクタを利用しても正しくされません。コンテンツを取得しようとするとエラーになり、ジョブは中断されます。

ファイルセキュリティを無効にすると、ジョブでクロールするすべてのコンテンツに索引アクセストークンを追加することができます。ただしこのトークンはサーバのトークンと一致している必要があります。トークンを入力して「追加」ボタンを押下してください。この機能はデモ以外に使われることは少ないと思います。

「メタデータ」タブを選択すると次のようなページが表示されます：

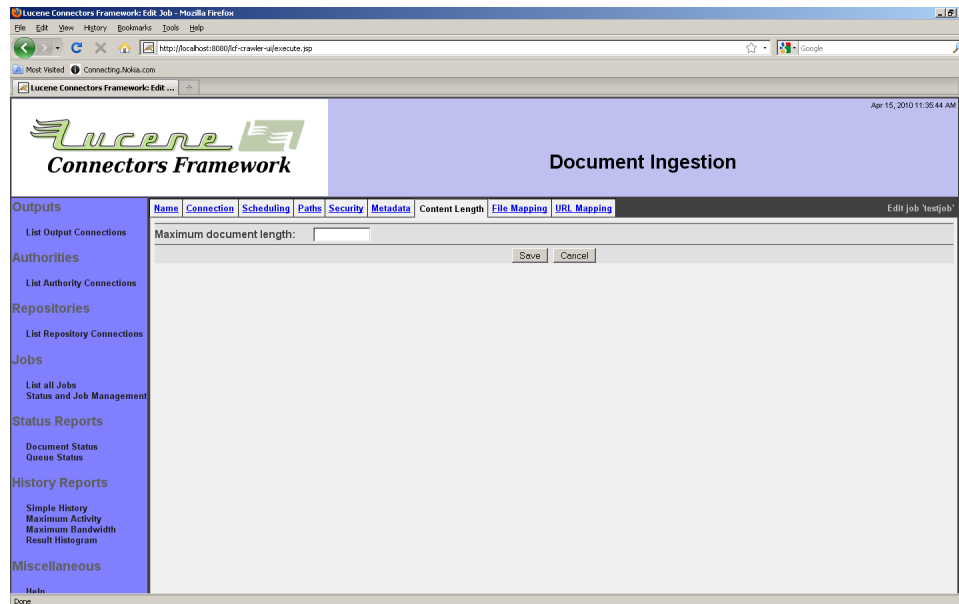


正規表現により変換されたコンテンツパスをコンテンツのメタデータとして取得できるようにする設定を行えます。「パス属性名」にメタデータ名を入力した後にルール一覧にルールを追加してください。各ルールは一致する正規表現の式で構成されます。変換元と値は括弧（「(」と「)」）で囲みます。括弧に囲まれた部分を「グループ」と言います。置き換え文字列は、固定文字と置き換えグループから構成されます。例えば、「\$(1)」は最初に一致したグループを示し、「\$(11)」は最初に一致した小文字のグループを示します。同じように「\$(1u)」は大文字にマップしたグループを示します。

例えば、ルール`http://(.*)/(.*)/`と置き換え文字列`http://$(2)/`は、`http://Server/Folder_1/Filename`を`http://Folder_1/Filename`に置き換えます。

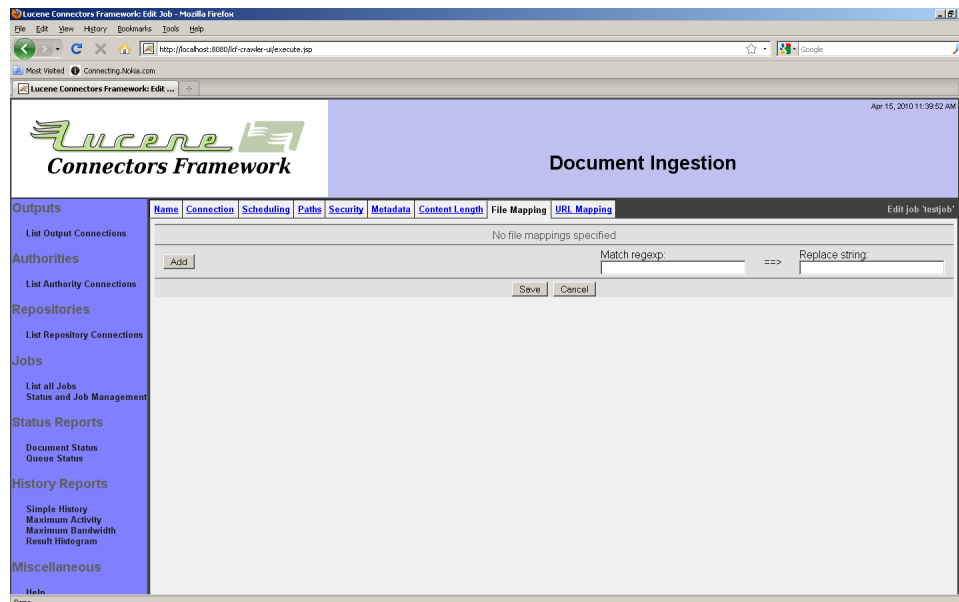
1つ以上のルールが存在する場合は、上から実行され、上のルールの結果は下のルールで変更されます。

「Content Length」タグを選択すると次のようなページが表示されます：



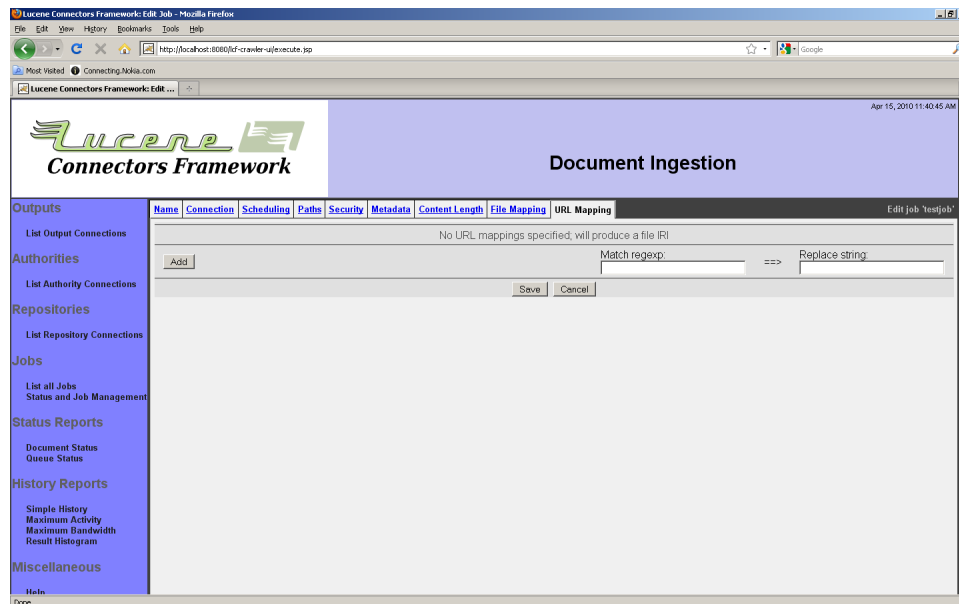
ジョブが長いコンテンツの作成を作成しないように、コンテンツを処理する最大長さを指定することができます。指定した長さより長いコンテンツは、指定した長さで切り捨てられます。最大長さを入力してください。最大長を設定しない場合は、空白にしてください。

「ファイルマップ」タブを選択すると次のようなページが表示されます：



ファイルマップはパス属性マップと同じように設定します。ファイルマップは実ファイルパスを変換します。元コンテンツと抽出したデータの間に変換が必要な場合に使うことができます。

「URLマップ」タブを選択すると次のようなページが表示されます：



URLマップはパス属性マップと同じように設定します。マップが存在しない場合は、ファイルパスは標準ファイルIRIに変換されます。マップが存在する場合は、Windows共有HTTPサーバを介して取得できる正しいURIへ変換されることを前提にします。

### 5.5 汎用データベースリポジトリコネクション

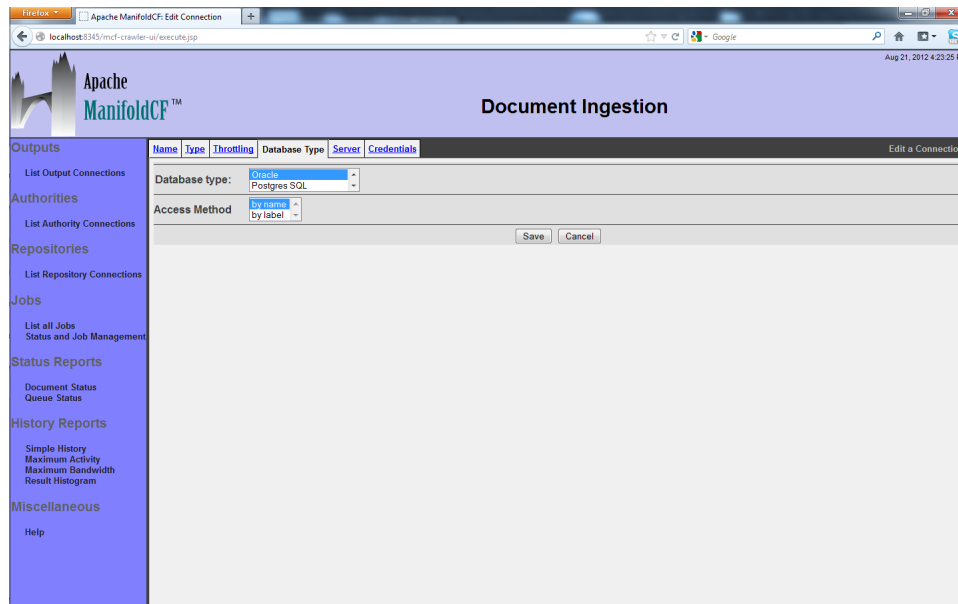
汎用データベースコネクションタイプは次のデータベースのテーブルの内容から索引を作成します：

- Postgresql (Postgresql JDBCドライバ)
- SQL Server (JTDS JDBCドライバ)
- Oracle (Oracle JDBCドライバ)
- Sybase (JTDS JDBCドライバ)
- MySQL (MySQL JDBCドライバ)

その他のデータベースを利用する場合はソフトウェアを修正する必要があります。データベースのセットアップによっては、利用できないデータベースもあります。

汎用データベースコネクションタイプはコンテンツ単位のセキュリティをサポートしていません。ただし、ジョブ単位ですべてのコンテンツのセキュリティを指定することはできます。設定するにはアクセストークンが必要になります。

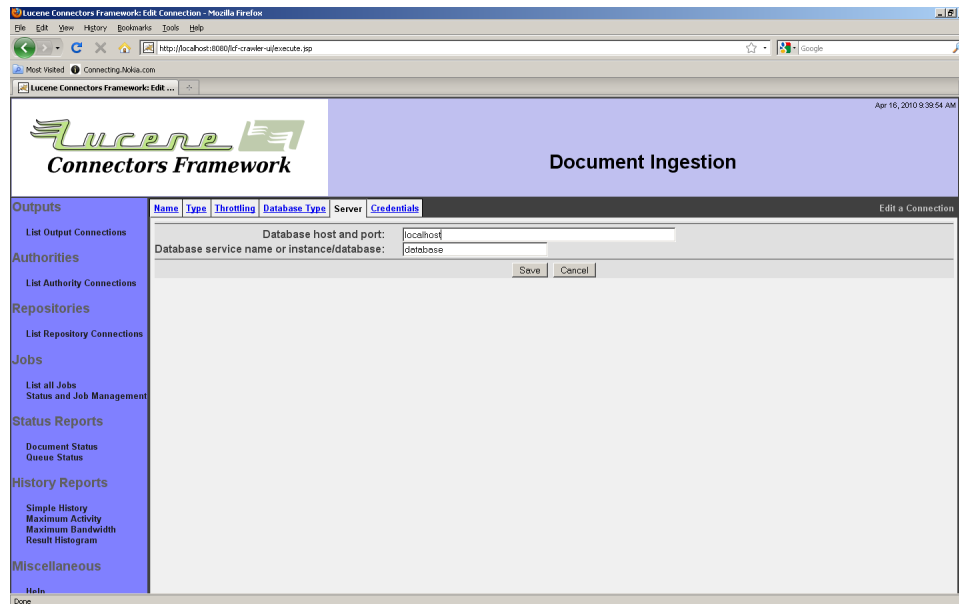
リポジトリコネクション編集ページから汎用データベースコネクションを選択すると3つのタブが表示されます：「データベースタイプ」、「サーバ」、「認証」。「データベースタイプ」タブを選択すると次のようなページが表示されます：



プルダウンメニューから索引を作成するデータベースの種類を選択してください。

また、JDBCアクセス方式をプルダウンから選択します。このアクセス方式は、JDBC仕様で最近明らかにされたのですが、カラム名の取得に関してすべてのJDBCドライバが同じ方法で動作するとは限らない、ということに基づいて提供されました。"by name"オプションは現在のところ、MySQLドライバを除いて、リスト内のすべてのJDBCドライバで動作します。"by label"は現在のMySQLドライバで動作します。他のドライバでも動作するかもしれません。汎用データベースのジョブで定義したクエリが正常動作しない場合や、カラムを見つけることができないというようなエラーメッセージが表示された場合は、このプルダウンを変更することによって解決するかもしれません。

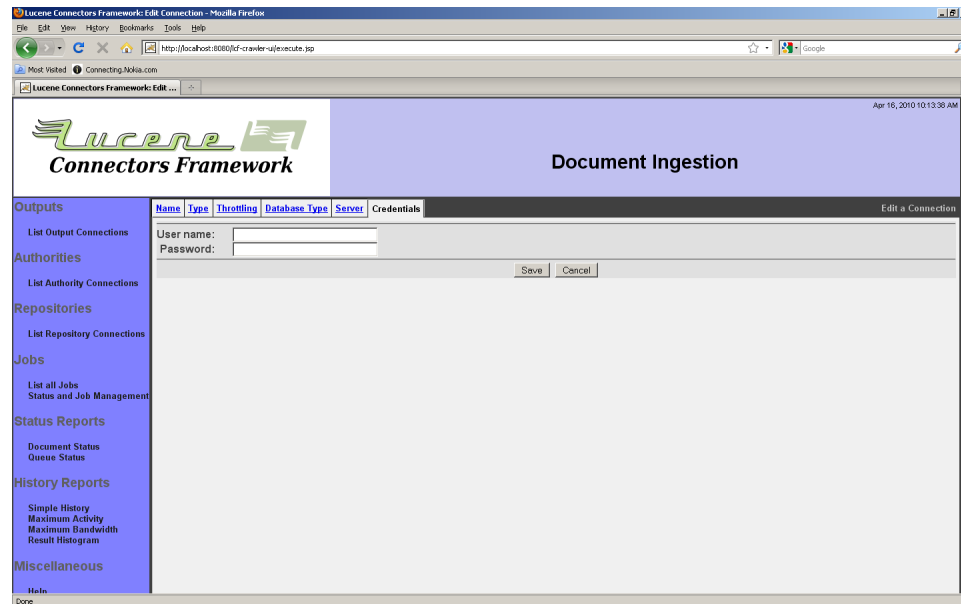
「サーバ」タブを選択すると次のようなページが表示されます：



サーバ名とポート番号を項目「データベースホストとポート」に入力してください。例えば、Oracleデータベースのデフォルトポート番号は1521なので、「my-oracle-server:1521」のように入力します。Postgresqlのデフォルトポート番号は5432なので、「my-postgresql-server:5432」のように入力します。SQL Serverのデフォルトポート番号は1433なので、「my-sql-server:1433」のように入力します。

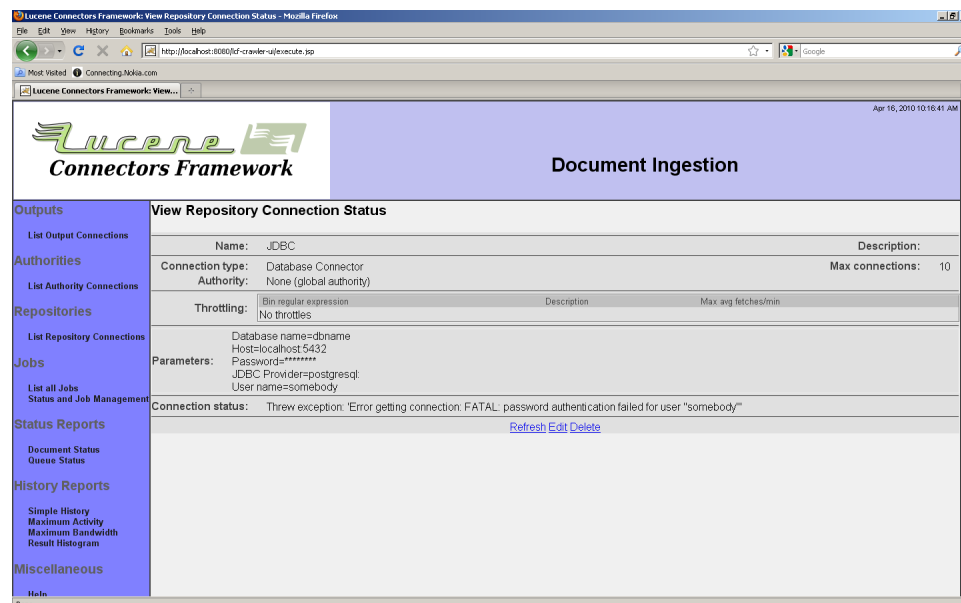
サービス名及びインスタンス名には利用するデータベースのインスタンス名を入力してください。Oracle及びPostgresqlの場合は、データベース名を入力してください。SQL Serverの場合は「my-instance-name/my-database-name」のように入力してください。SQL Serverのデフォルトのインスタンスを利用する場合は、データベースのみを入力してください。

「認証」タブを選択すると次のようなページが表示されます：



データベースに接続するためのユーザ認証情報を入力してください。

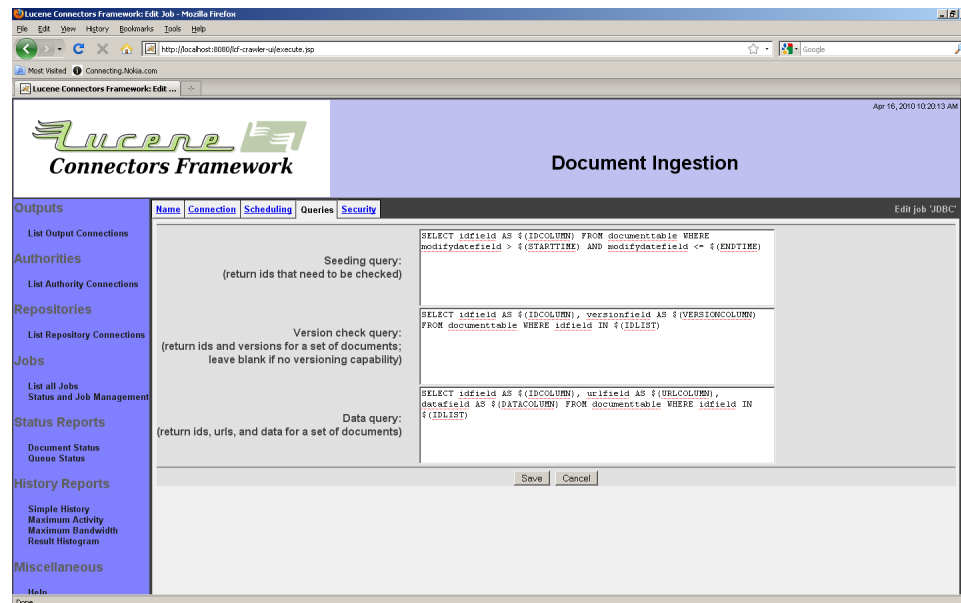
入力した後に「保存」ボタンを押下してください。次のような設定内容概要ページが表示されます：



画面例では汎用データベースコネクションはデータベースと接続できないためにエラーメッセージが表示されています。

ジョブのリポジトリコネクションに汎用データベースコネクションタイプを選択すると「クエリー」と「セキュリティ」タブが表示されます。

「クエリー」タブを選択すると次のようなページが表示されます：



最低でも2つのクエリーを設定する必要があります(3つめのクエリーは任意です)。これらのクエリーに基づいてデータベースからコンテンツを取得します。クエリーを入力する前に、データベース構造をどのようにManifoldCFフレームワークの構造に対応付けるか決める必要があります。

- 時間帯に発生する追加及び変更したコンテンツID一覧を取得します(下を参照)。
- コンテンツIDからバージョン情報を取得します(下を参照)。
- コンテンツIDとバージョン情報からコンテンツ情報を取得します。コンテンツ情報は、コンテンツの内容、URI、メタデータから構成されます。

ManifoldCFはリポジトリコネクション内のすべてのコンテンツを識別するためにIDを使います。このIDはコンテンツを取得するために主キーとして使われます。ジョブの汎用データベースタイプで使うデータベースにも同じような概念が必要です。間違ったIDを選択した場合は、処理が遅くなる場合もあります。

時間帯に発生したすべての変更のコンテンツID一覧を取得する必要があります。変更されていないコンテンツのIDを一覧に含めることも出来ますが、最適ではありません。

コンテンツを追加するように設定する場合は、「バージョン情報」も設定してください。この文字列を利用してコンテンツが変更したかを確認します。索引を変更する場合は、このバージョン情報も変わる必要があります。(その他の原因で変わっても問題はありません。)



入力されたクエリーはコネクションが使う前にクエリーに含まれている置き換え文字列に値が入れられます。デフォルトで入力されているクエリーには代表的な置き換え文字列が利用されています。例えば、「\$(IDCOLUMN)」はコネクションが利用するIDの列名に置き換えられます。その他の置き換え文字列は次の通りです：

名前	説明
IDCOLUMN	コンテンツIDを含む結果セットの列
VERSIONCOLUMN	バージョン情報を含む結果セットの列
URLCOLUMN	URIを含む結果セットの列
DATACOLUMN	コンテンツデータを含む結果セットの列
STARTTIME	開始時間を1970年1月1日からの経過時間(ミリ秒)
ENDTIME	終了時間を1970年1月1日からの経過時間(ミリ秒)
IDLIST	括弧で囲まれたコンテンツID一覧

時間の置き換え文字列を含むクエリーを作成する場合は、「\$(STARTTIME)」と「\$(ENDTIME)」は1970年1月1日からの経過時間をミリ秒で表した値に置き換えられることに注意してください。「\$(STARTTIME)」と「\$(ENDTIME)」をシステムのタイムスタンプに置き換えることを推奨します。

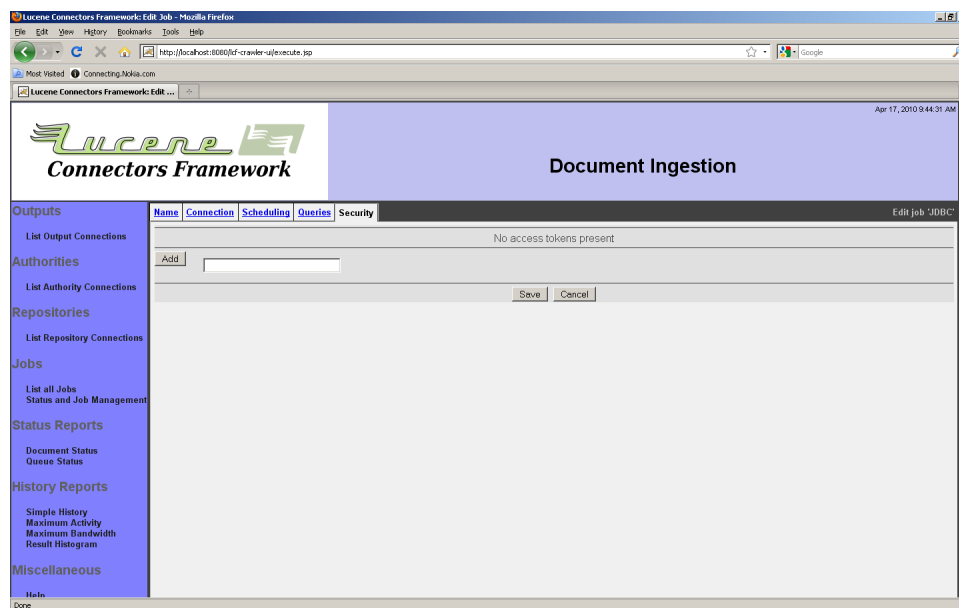
以下は「\$(STARTTIME)」と「\$(ENDTIME)」を他形式の日付と時間に変換するサンプルクエリーの部分です。先頭の列はクエリー句が対応するSQLデータベースです。2列目はクエリーの出力データ型です。3列目は、「\$(STARTTIME)」を利用した例です。これらはクエリーの例です。変更したデータ型が2列目のデータ型と異なる場合は、日付比較は正しくされない場合もあります。

データベース	型	サンプルクエリー
Oracle	date	TO_DATE ( '1970/01/01:00:00:00', 'yyyy/mm/ dd:hh:mi:ss') + ROUND (\$ (STARTTIME)/86400000)
Oracle	timestamp	TO_TIMESTAMP('1970-01-01 00:00:00') + interval '\$(STARTTIME)/1000' second
Postgres SQL	timestamp	date '1970-01-01' + interval '\$(STARTTIME) milliseconds'

MS SQL Server (\$>\$6.5)	datetime	DATEADD(ms, \$(STARTTIME), '19700101')
Sybase (10+)	datetime	DATEADD(ms, \$(STARTTIME), '19700101')

汎用データコネクションを利用したジョブを作成する場合は、ジョブのクエリーには例題クエリーが入力されています。クエリーが返す列はこれらを参考にして下さい。多くの場合は、例題で返されている列のみで充分です。ただし、ファイルデータクエリーの場合はそれ以外の列も返すことがあります。この場合は、列値はコンテンツのメタデータとして索引に渡されます。メタデータ名は結果セットの列名になります。

「セキュリティ」タブは、汎用データベースジョブで作成された索引のコンテンツにアクセストークンを追加します。追加するトークンは、どの権限コネクションを利用するか決めてその権限コネクションのアクセストークンに依存します。「セキュリティ」タブを選択すると次のようなページが表示されます：



アクセストークンを選択して、「追加」ボタンを選択してください。複数のアクセストークンを設定することも可能です。

## 5.6 IBM FileNet P8リポジトリコネクション

IBM FileNet P8コネクションタイプは、FileNet P8サーバーインスタンスのコンテンツから索引を作成する場合に利用します。1つのコネクションからそのサーバにあるファイルすべてを取得できます。複数のサーバーをサポートするには複数のコネクションが必要です。

このコネクションタイプはActive Directory権限コネクションを使ってコンテンツの権限を管理します。まだActive Directory権限コネクションを作成しておらず、コンテンツの権限を管理したいのであれば、「Active Directory権限コネクション」の節の指示に従ってください。

FileNetを選択すると次の独自のタブが表示されます:「サーバ」、「オブジェクトストア」、「ドキュメントURL」、「認証」。「サーバ」タブでは接続するFileNet P8の指定が、「オブジェクトストア」タブでは希望するFileNetオブジェクトストアの指定ができます。「ドキュメントURL」タブでは索引を作成するコンテンツの各URLのパラメータを、「認証」タブではFileNetオブジェクトストアへのアクセスに使う認証情報を指定します。

「サーバ」タブを選択すると次のようなページが表示されます:

名前	タイプ	スロットリング	サーバ	オブジェクトストア	ドキュメントURL	認証	コネクションを編集
出力							
変換コネクション一覧							
出力コネクション一覧							
権限							
権限グループ一覧							
ユーザーマッピングコネクション一覧							
権限コネクション一覧							
リポトリ							

サーバプロトコル:   
 サーバプロトコル:   
 サーバホスト名:   
 サーバポート:   
 サーバWebサービス:

適切なプロトコルを選択し、サーバ名、ポート、サービスを指定してください。

「オブジェクトストア」タブを選択すると次のようなページが表示されます:

名前	タイプ	スロットリング	サーバ	オブジェクトストア	ドキュメントURL	認証	コネクションを編集
出力							
変換コネクション一覧							
出力コネクション一覧							
権限							
権限グループ一覧							
ユーザーマッピングコネクション一覧							

FileNetドメイン名:   
 オブジェクトストア名:

接続するFileNetドメイン名とドメイン内のFileNetオブジェクトストア名を入力してください。

「ドキュメントURL」を選択すると次のようなページが表示されます:

名前	タイプ	スロットリング	サーバ	オブジェクトストア	ドキュメントURL	認証	コネクションを編集
出力							
変換コネクション一覧							
出力コネクション一覧							
権限							
権限グループ一覧							
ユーザーマッピングコネクション一覧							
権限コネクション一覧							
リポトリ							

ドキュメントURLプロトコル:   
 ドキュメントURLプロトコル:   
 ドキュメントURLホスト名:   
 ドキュメントURLポート:   
 ドキュメントURL:

このタブでは索引を作成するコンテンツをどのようにロードするか of の基盤となるURLを指定できます。プロトコルを選択してください。ホスト名を入力してください。ポートを入力してください。そして、URLを入力してください。

「認証」タブを選択すると次のようなページが表示されます:

出力	名前	タイプ	スロットリング	サーバ	オブジェクトストア	ドキュメントURL	認証	接続を編集
変換コネクション一覧								
出力コネクション一覧	ユーザID: <input type="text"/>							
権限	パスワード: <input type="text"/>							
権限グループ一覧								
ユーザーマッピングコネクション一覧	<input type="button" value="保存"/> <input type="button" value="キャンセル"/>							

FileNetリポジトリに対してFileNetコネクションタイプアクセスできるようにFileNetユーザーIDとパスワードを入力してください。

コネクション情報の入力が終わったら、「保存」ボタンを押してください。次のようなページが表示されます:

出力	リポジトリコネクション状態の表示
変換コネクション一覧	名前: test 説明:
出力コネクション一覧	コネクションタイプ: FileNet 最大コネクション数: 10
権限	権限グループ: なし (globalAuthority)
権限グループ一覧	スロットリング: 正解表無 説明 最大平均取得/分
ユーザーマッピングコネクション一覧	スロットリングなし
権限コネクション一覧	Server WebServices location=ws/FNCEWS40DIME Document URL hostname=localhost Server hostname=localhost Password=***** Document URL location=Workplace/Browse.jsp Document URL port=2345 Server port=1234 Object store=bar FileNet domain=foo User ID=me Document URL protocol=http Server protocol=http
リポジトリ	コネクションの状態: Connection temporarily failed: Connection refused to host: 127.0.0.1; nested exception is: java.net.ConnectException: Connection refused
リポジトリコネクション一覧	<input type="button" value="更新"/> <input type="button" value="編集"/> <input type="button" value="削除"/> <input type="button" value="すべての関連履歴をクリア"/>
List Notification Connections	
ジョブ	
ジョブ一覧	
状態とジョブ管理	
状態レポート	
コンテンツの状態	

## 5.7 EMC Documentumリポジトリコネクション

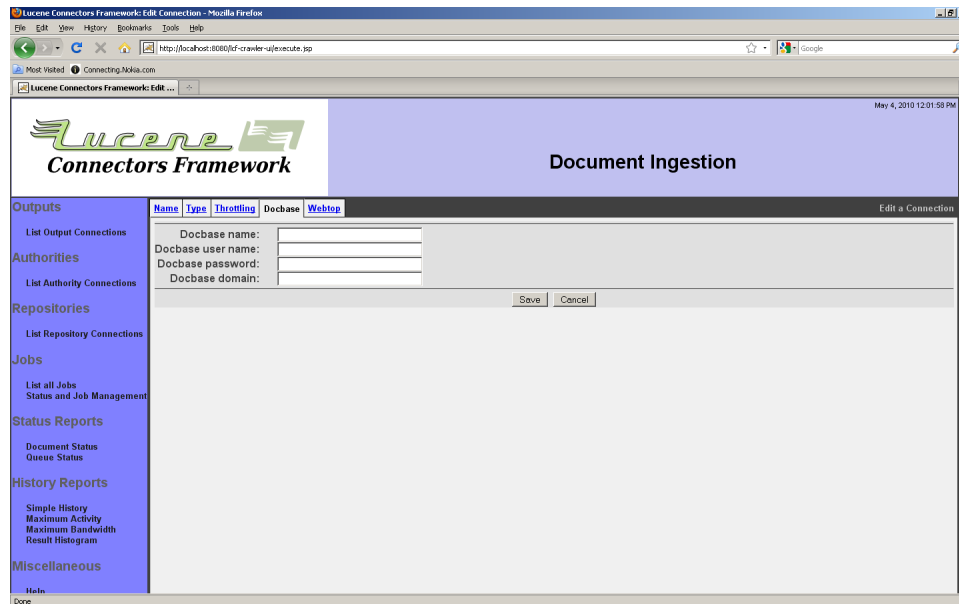
EMC Documentumコネクションタイプは、Documentum Content Serverインスタンスのコンテンツから索引を作成する場合に利用します。1つのコネクションから1つのContent Serverインスタンスのすべてのコンテンツを取得することができます。複数のContent Serverインスタンスのコンテンツの索引を作成する場合は、サーバインスタンス毎にコネクションを定義する必要があります。

Documentumコネクションタイプは、Content Serverインスタンス毎にdm\_document型及びdm\_document型から派生した型のすべてのDocumentumコンテンツから索引を作成できます。複合ドキュメントは構成する複合ドキュメントに対応しています。その他のDocumentum構成には未対応です。

Documentumコネクションで扱うコンテンツはDocumentum権限コネクションで権限管理されています。Documentum権限を利用する場合は「EMC Documentum権限コネクション」を参照してください。

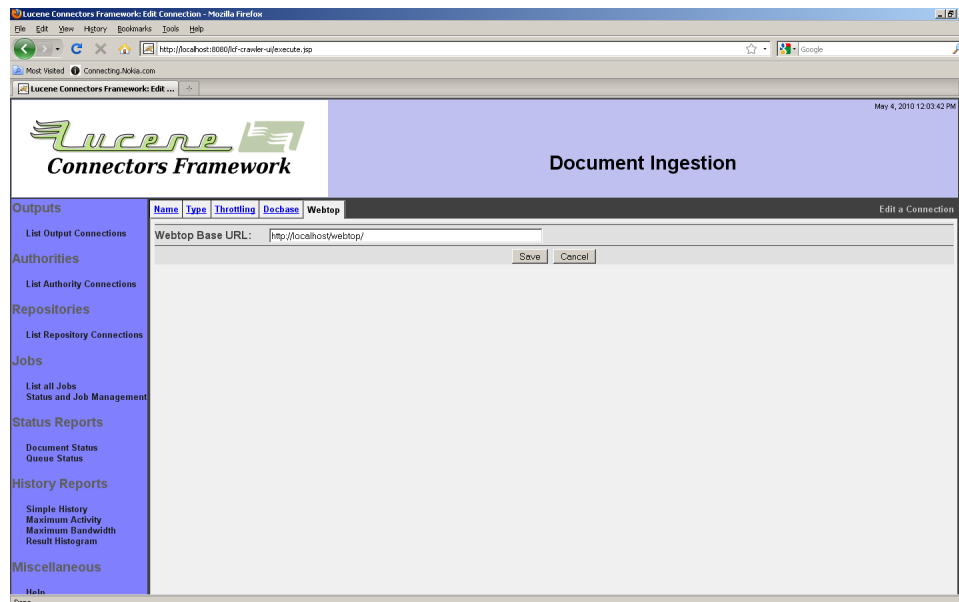
Documentumコネクションを選択すると次のタブが表示されます:「Docbase」、「Webtop」。「Docbase」タブからは接続するコンテンツサーバの指定と、認証情報の設定ができます。索引を作成した後に、「Webtop」タブからはコンテンツサーバの内容を表示するWebtopサーバの指定することができます。

「Docbase」タブを選択すると次のようなページが表示されます:



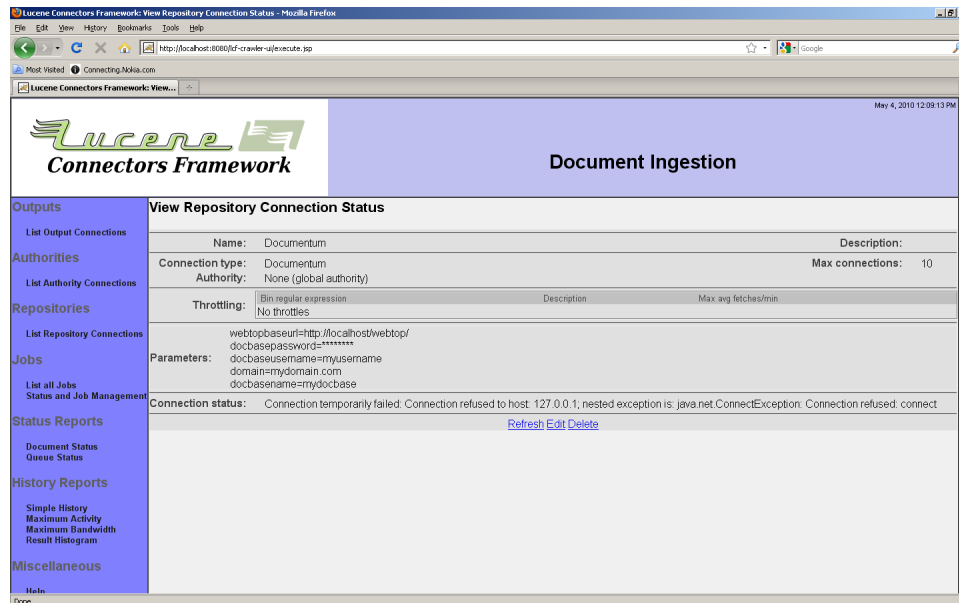
コンテンツサーバDocbaseインスタンス名と認証情報を入力してください。コンテンツサーバインスタンスがActive Directoryと統合されていない場合は、「ドメイン」を空白にしてください。

「Webtop」タブを選択すると次のようなページが表示されます；



WebtopインスタンスのベースURIを入力してください。このURIはユーザが元コンテンツを参照する場合のみ利用されます。クロールには利用されません。

入力をした後に「保存」ボタンを押下してください。次のような設定内容の概要と状態が表示されます：



状態にエラーメッセージが表示されている場合は、設定内容を修正してください。

ジョブでDocumentumコネクションを選択すると、次のタブが表示されます：「パス」、「ドキュメントタイプ」、「コンテンツタイプ」、「セキュリティ」、「パスメタデータ」。

「パス」タブからはコンテンツを読み込むDocumentum内のパスを設定することができます。パスが設定されていない場合は、すべてのコンテンツを読み込みの対象にします。

「ドキュメントタイプ」タブからは、読み込みの対象にするドキュメントタイプを指定します。システム管理者が索引対象と指定したdm\_documentから派生したドキュメントタイプのみから選択することができます。また、索引を作成するドキュメントタイプ毎に含むメタデータを指定することもできます。ドキュメントタイプのすべてのメタデータを含む場合は「すべてのメタデータ」チェックボックスをチェックしてください。

「コンテンツタイプ」タブからは、コンテンツセットに含むDocumentumのmimeタイプを指定することができます。含むタイプにチェックをし、除外するタイプからはチェックを外してください。

「セキュリティ」タブからは、このジョブでDocumentumセキュリティを有効／無効にするか指定することができます。Documentumセキュリティを無効にする場合は「無効」ラジオボタンを選択してください。無効にした場合は、ジョブのすべてのコンテンツ取得で利用するアクセストークンを設定することができます。アクセストークンは利用する権限コネクションタイプによります。アクセストークンの入力毎に「追加」ボタンを押下してください。

「パスメタデータ」タブからはコンテンツ毎のパス情報を索引にメタデータとして送るように指定することができます。送るようにする場合は、項目「パス属性名」にメタデータ属性名を入力して、ルールをルール一覧に追加してください。各ルールに一致する正規表現の式で構成されます。変換元と値は括弧（「(」と「)」）で囲みます。括弧に囲まれた部分を「グループ」と言います。置き換え文字列は、固定文字と置き換えグループから構成されます。例えば、「\$(1)」は最初に一致したグループを示し、「\$(1l)」は最初に一致した小文字のグループを示します。同じように「\$(1u)」は大文字にマップしたグループを示します。

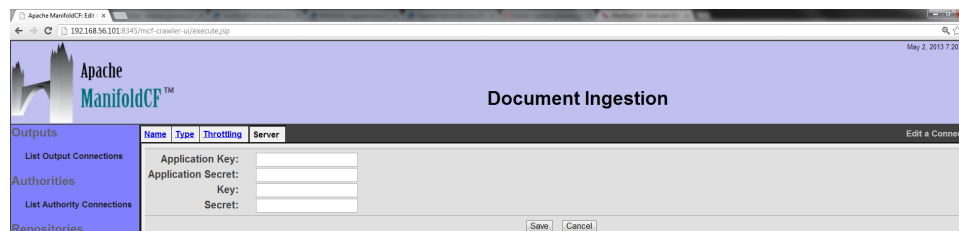
例えば、ルールの一致条件が`*/(.*)/(.*)/*`で置き換え文字列`$(1) $(2)/`とした場合、パス`Project/Folder_1/Folder_2/Filename`は`Folder_1 Folder_2`に変換されます。

1つ以上のルールが存在する場合は、上から実行され、上のルールの結果は下のルールで変更されます。

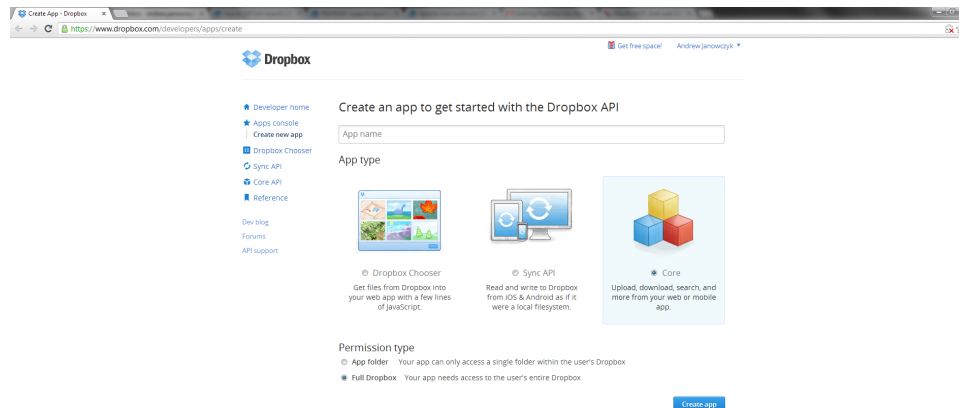
## 5.8 Dropboxリポジトリコネクション

Dropboxリポジトリコネクションは、Dropboxの内容をインデクシングすることができます。それぞれのDropboxコネクションは、ひとつのDropboxリポジトリへのアクセスを管理します。これは、たとえば異なるユーザを使って、複数のDropboxを持っている場合、それぞれのDropboxリポジトリに対してひとつずつコネクションを作り、その関連した権限情報を用意する必要があることを意味します。

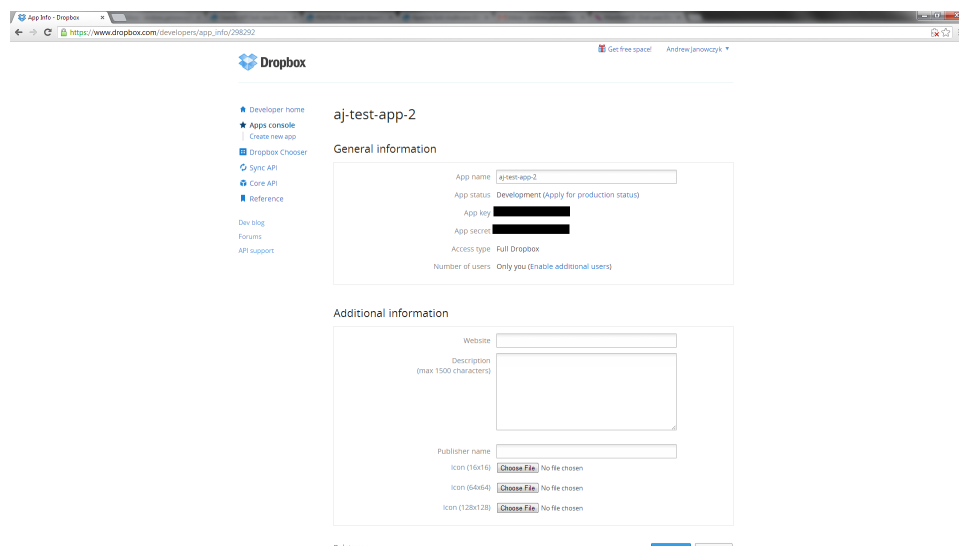
ひとつのDropboxコネクションは、次のような設定パラメータを、リポジトリコネクションの編集画面に持っています。



コネクションに接続するためには、4つの情報が必要です。Application KeyとApplication Secretは、開発ライセンスであなたのアプリケーションを登録した時に、Dropboxから提供されます。これは基本的には、アプリケーション開発者用のDropbox websiteを通して行われます。



今回の用途としては、DropboxとコミュニケーションするためにRESTサービスを使いますので、アプリケーションタイプとして"Core"を選択する必要があります。また、"full access"を選択します。これには少々議論があります。基本的に、情報を格納したり取得したりするアプリケーションでは、アプリケーションの固有フォルダからフルアクセスします。今回のケースでは、ユーザーがユーザーのファイルをそのままアクセスし、manifoldcfの固有フォルダにコピーしないことを想定しています。結果的には、"App folder"の代わりに"full access"を選択しています。



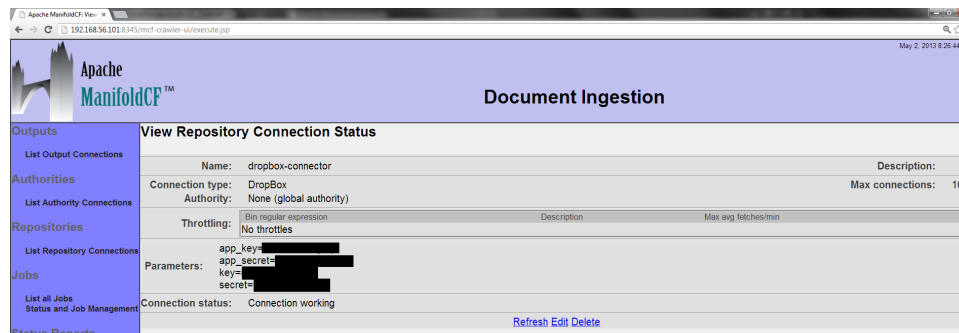
その後、このコネクタで求められる2つの情報、App keyとApp secretを見ることができます。

ここで、それぞれのユーザは、ユーザのアプリケーションがDropboxにアクセスできるよう受諾されることを確認しなければなりません。これは普通のOAUTHアプローチを通してなされます。ユーザのアプリケーションのkeyとsecretが提供された後、ユーザは、ユーザのアプリ

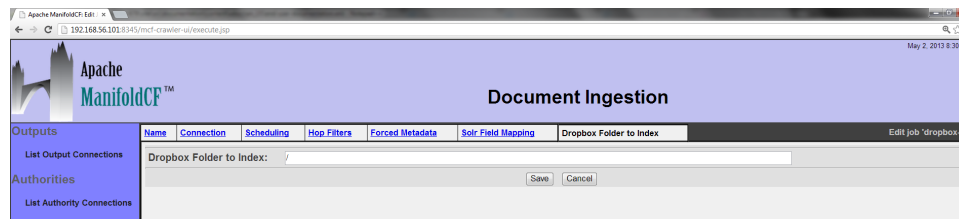


リケーションの権限を許可してもらうことを、Dropboxのウェブサイトに対して問い合わせるよう指示されます。彼らがそのリクエストを受諾すると、Dropboxはclient keyとsecretを提供します。このkeyとsecretが、Dropboxコネクタで必要になる最後の2つです。このプロセスの深い内容については、dropbox websiteで説明されており、どのようにこの2つのclientトークンを生成するか例が示されています。

保存ボタンをクリックしたら、コネクションのサマリ画面を見ることになります。これは以下のようなものになります。



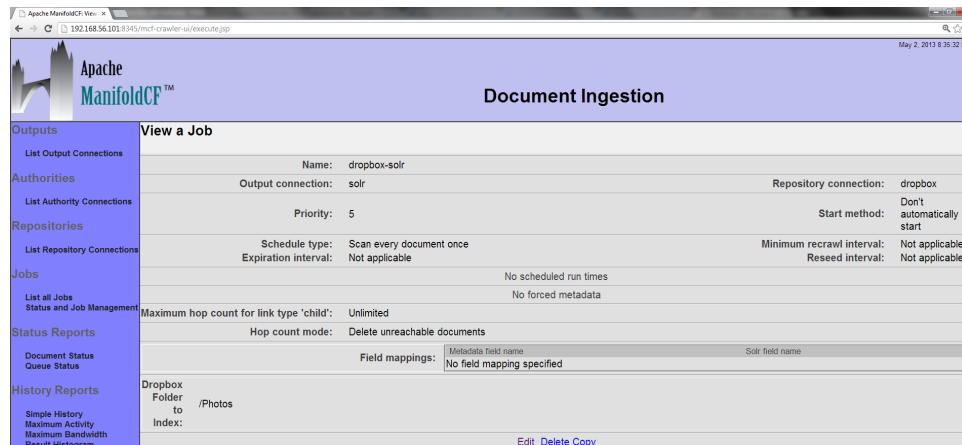
Dropboxリポジトリコネクションを使用するジョブを設定した場合は、追加タブが表示されます。これは、"Dropbox Folder to Index" です。



このタブでは、Dropboxコネクタがインデクシングするディレクトリを指定することができます。Dropboxはunixスタイルのパスを使います。"/"はルートパスを意味します(したがって全体のDropboxを指定することになります)。たとえば、Photosディレクトリをインデクシングしたい場合は、"/Photos"と指定することになります。

注意点は、Dropboxコネクタは取り込み処理中、それぞれの結果について、フォルダの階層を見つけた時、そのフォルダの子供のすべてのフォルダを取り込もうとします。フォルダでなければ、直接ドキュメントを取り込もうとします。

ジョブの設定が終わったら、保存ボタンをクリックし、サマリ画面を見ます。これは以下のようなものになります。



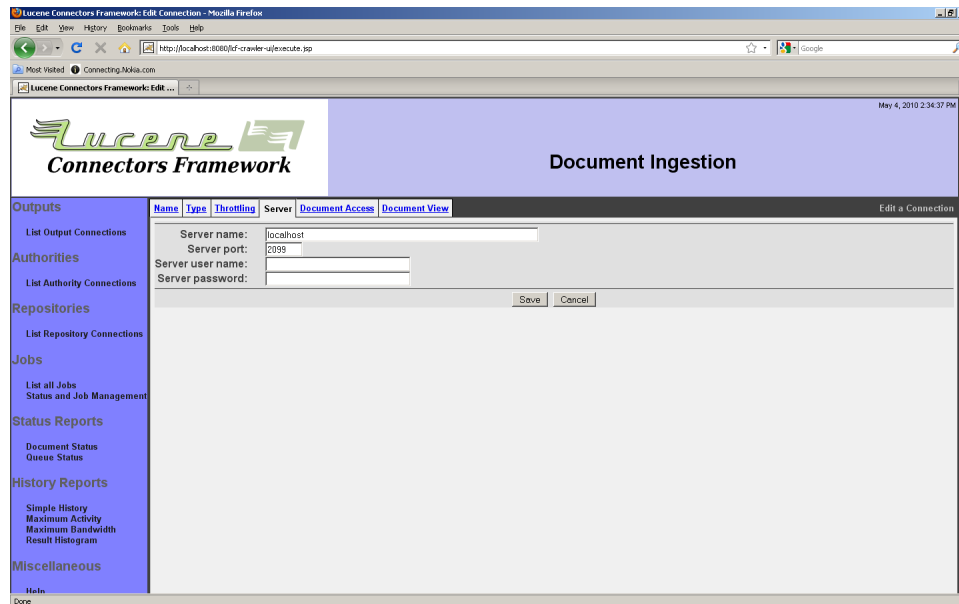
## 5.9 OpenText LiveLinkリポジトリコネクション

OpenText LiveLinkコネクションタイプは、LiveLinkリポジトリからのコンテンツから索引を作成します。LiveLinkには基本ドキュメント、複合ドキュメント、フォルダ、ワークスペース、プロジェクトのような多くのドキュメントタイプがあります。LiveLinkコネクションはこれらのすべてのドキュメント種類のコンテンツを処理することができます。

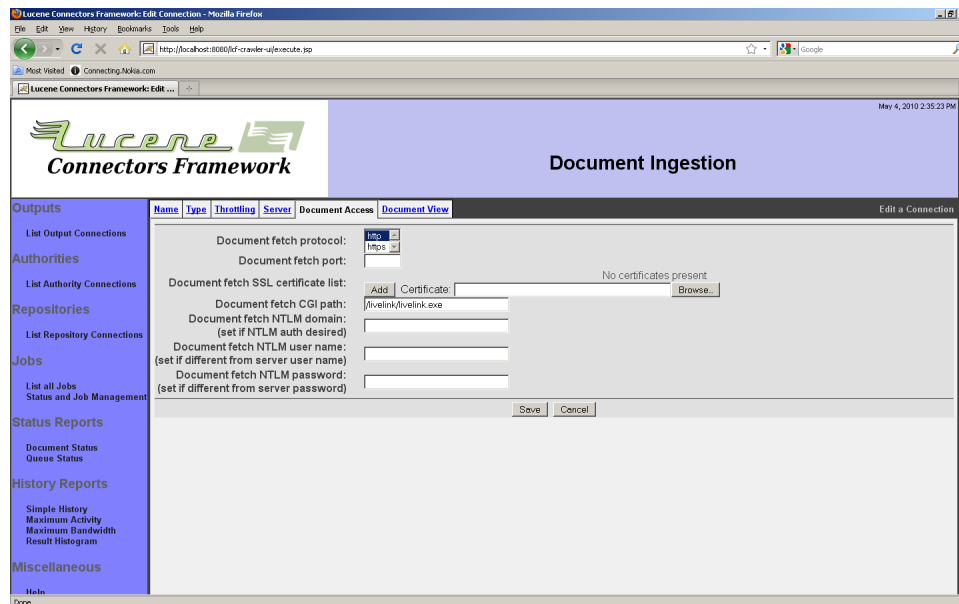
LiveLinkコネクションで処理するコンテンツのセキュリティはLiveLink権限で管理されています。LiveLink権限コネクションについては「OpenText LiveLink権限コネクション」を参照してください。

LiveLinkコネクションを選択すると次のタブが表示します:「サーバ」、「ドキュメントアクセス」、「ドキュメント表示」。「サーバ」タブからは、接続するLiveLinkサーバの選択と接続するための認証情報を設定することができます。「ドキュメントアクセス」タブからはLiveLinkからコンテンツを取得するためのLiveLinkのwebインタフェースの情報を設定します。「ドキュメント表示」タブからは、検索結果を表示する取得したコンテンツのURIを作成方法を指定します。

「サーバ」タブを選択すると次のようなページが表示されます:



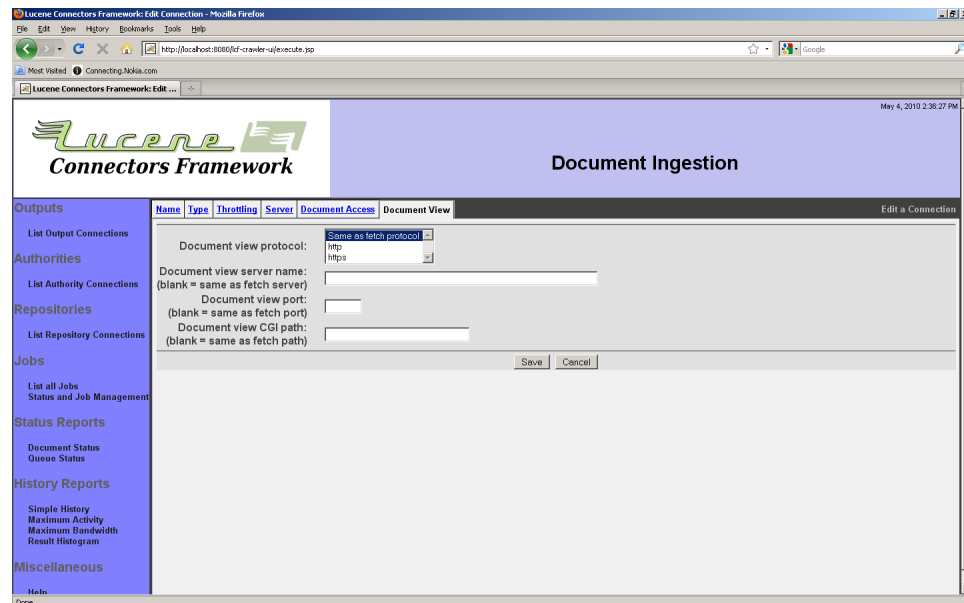
LiveLinkサーバ名、ポート番号、認証情報を入力してください。  
 「ドキュメントアクセス」タブを選択すると次のようなページが表示されます：



サーバ名は「サーバ」タブに入力した値と同じと想定されます。プロトコルを選択してください。LiveLinkサーバはHTTPデフォルト以外のポート番号を利用している場合はポート番号を入力してください。LiveLinkサーバがNTLM認証を利用している場合は、Active Directoryのユーザ名、パスワードとドメイン名を入力してください。LiveLinkサーバがHTTPSを利用している場合は、「追加」ボタンを押下して証明書をアップロードしてください

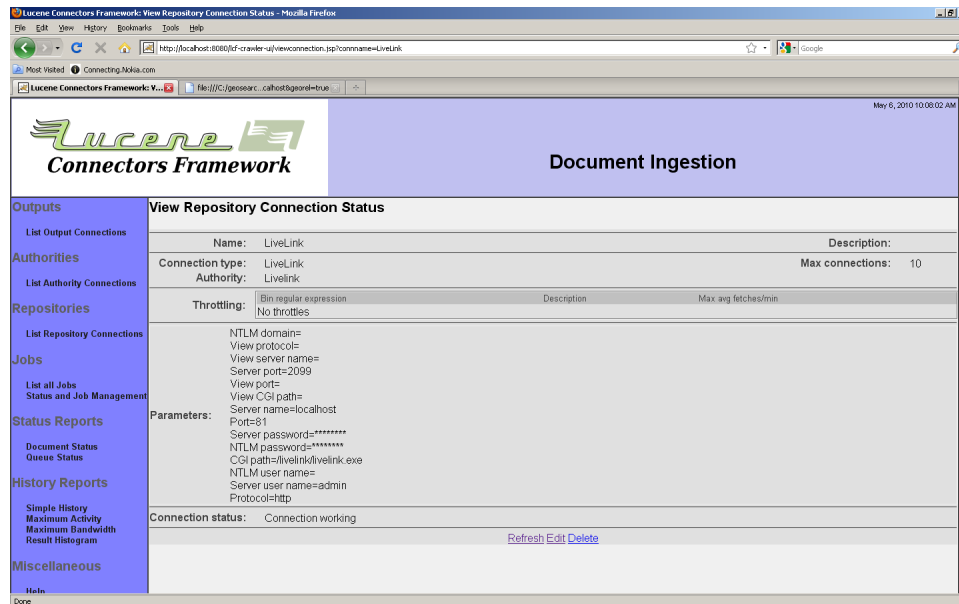
い。(サーバの証明書を使うこともできますが、サーバの証明書は変わる可能性がありますので注意してください。)

「ドキュメント表示」タブを選択すると次のようなページが表示されます:



各ドキュメントの表示URIをアクセスURIと同じにする場合は、内容を変更しないでください。検索結果を異なるCGIで表示する場合は、このページから設定を行ってください。

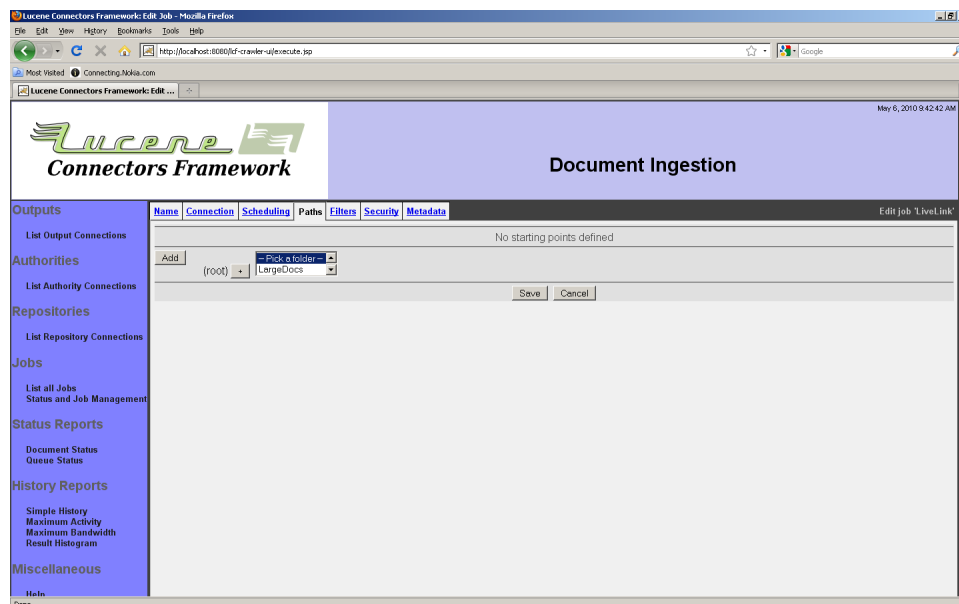
入力した後に「保存」ボタンを押下してください。次のような設定内容概要ページが表示されます:



エラーメッセージが表示された場合は、設定を修正してください。画面例では正しく設定されたため、コネクション状態は「起動」と表示されています。

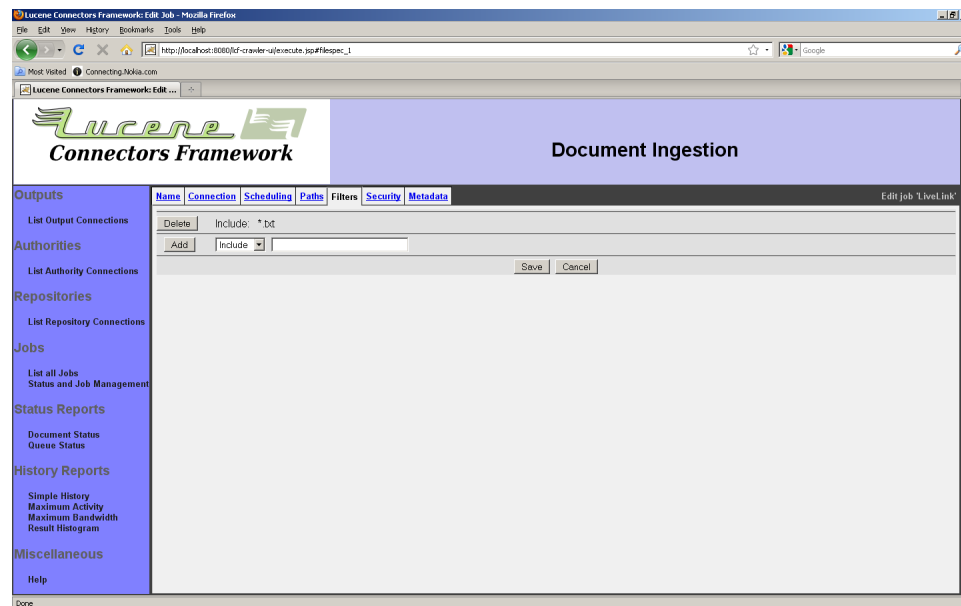
ジョブでLiveLinkコネクションを選択すると次のタブが表示します:「パス」、「フィルタ」、「セキュリティ」、「メタデータ」。

「パス」タブからはLiveLinkが索引を作成する起点となるパス一覧を設定します。「パス」タブを選択すると次のようなページが表示されます:



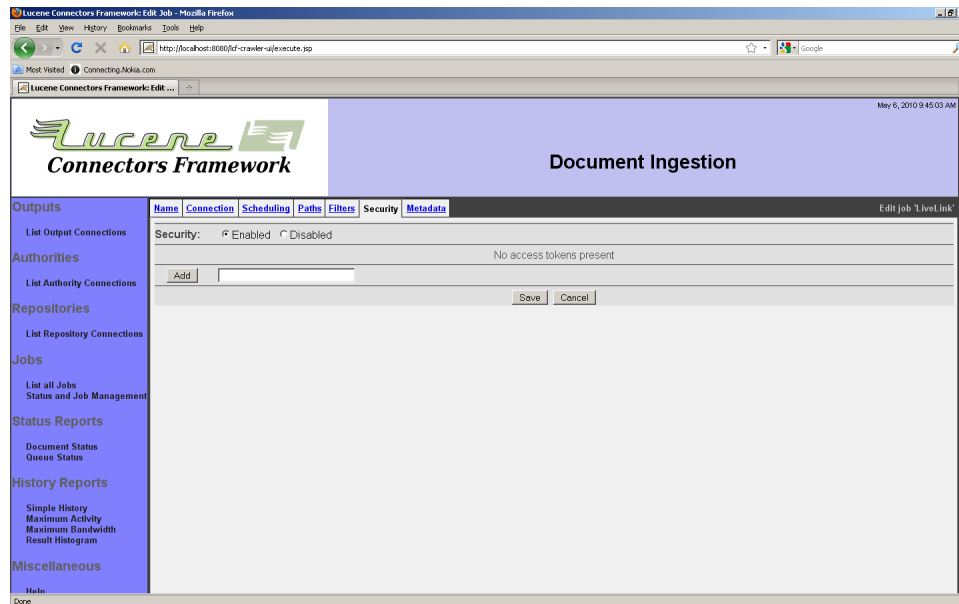
ドロップダウンリストから選択して「+」ボタンを押下してパスを作成してください。パスを作成し終わったら「追加」ボタンを押下して一覧に追加してください。

「フィルタ」タブからはLiveLinkジョブが対象する含む／除外するコンテンツの条件を指定します。ファイルはルール一覧です。各ルールはパスを特定する条件を指定します。各ルールはファイル名条件(例:「\*.txt」)、ファイルまたはフォルダ名を特定するのかの指定、出力コネクションでファイルから索引を作成するか否か、ファイルを含むか除外するかの指定が含まれます。ファイル名の指定にはワイルドカード文字「\*」と「?」を使うこともできます。「\*」は0以上の任意の文字と一致します。「?」は任意の1文字と一致します。その他の文字は記述通りに一致する必要があります。

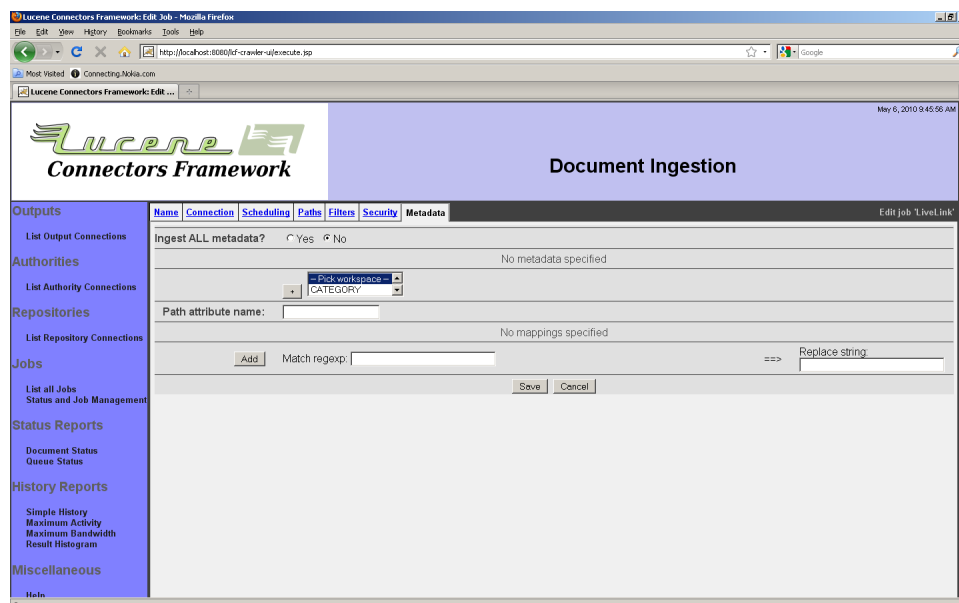


一致させる値を入力し、一致した場合の処理を選択した後に「追加」ボタンを押下してルールをフィルター一覧に追加してください。

「セキュリティ」タブからジョブで対象とするコンテンツのLiveLinkセキュリティを無効／有効に設定することができます。「セキュリティ」タブを選択すると次のようなページが表示されます:



セキュリティを無効にするとアクセストークンをジョブのすべての索引を作成するコンテンツに指定することができます。アクセストークンの形式はジョブのリポジトリコネクションで利用する権限によります。トークンを入力して「追加」ボタンを押下して一覧に追加してください。「メタデータ」タブからは索引に渡すLiveLinkのメタデータを指定することができます。「メタデータ」タブを選択すると次のようなページが表示されます：



LiveLinkのすべてのメタデータを索引に渡す場合は、「すべてのメタデータ」ラジオボタンをチェックしてください。特定のメタデータのみを渡す場合は、LiveLinkメタデータパスをメタ

データ一覧に追加してください。次のメタデータ句を選択して「+」ボタンを押下してパスに追加してください。フォルダ情報、メタデータカテゴリを追加することができます。

メタデータカテゴリに辿りついたら、メタデータ属性を選択するか、「このカテゴリのすべての属性」チェックボックスをチェックしてください。入力が終わったら「追加」ボタンを押下して索引に含むメタデータ属性を追加してください。

「パスメタデータ」タブからはコンテンツ毎のパス情報を索引にメタデータとして送るよう指定することができます。送るようになる場合は、項目「パス属性名」にメタデータ属性名を入力して、ルールをルール一覧に追加してください。各ルールに一致する正規表現の式で構成されます。変換元と値は括弧（「(」と「)」）で囲みます。括弧に囲まれた部分を「グループ」と言います。置き換え文字列は、固定文字と置き換えグループから構成されます。例えば、「\$(1)」は最初に一致したグループを示し、「\$(1l)」は最初に一致した小文字のグループを示します。同じように「\$(1u)」は大文字にマップしたグループを示します。

例えば、ルールの一致条件が`*/(.*)/(.*)/*`で置き換え文字列`$(1) $(2)/`とした場合、パス`Project/Folder_1/Folder_2/Filename`は`Folder_1 Folder_2`に変換されます。

1つ以上のルールが存在する場合は、上から実行され、上のルールの結果は下のルールで変更されます。

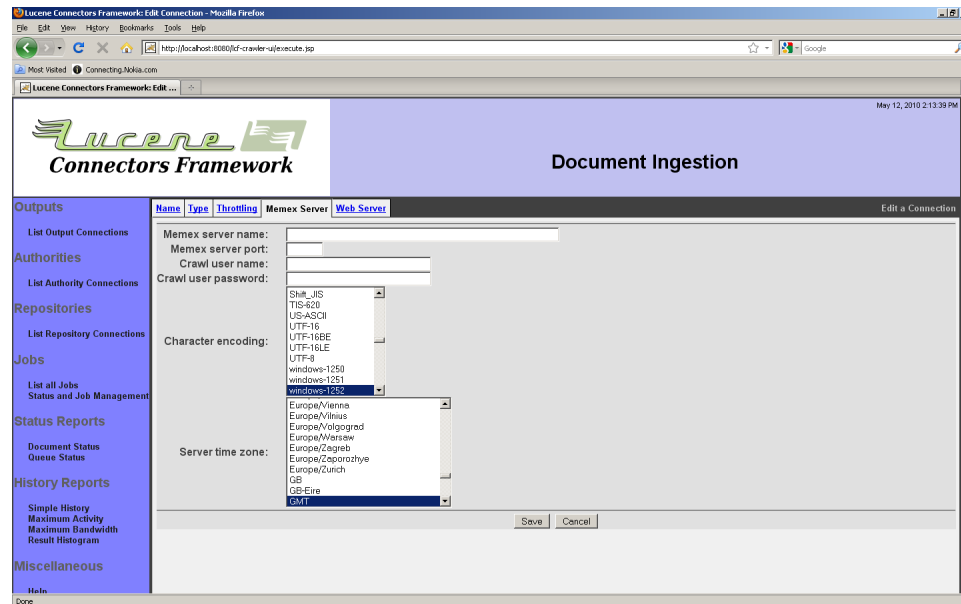
## 5.10 Memex Patriarchリポジトリコネクション

Memex PatriarchコネクションはMemexサーバのコンテンツの索引を作成します。

Memexコネクションで処理するコンテンツのセキュリティはMemex権限で管理されています。Memex権限コネクションの設定については「Memex Patriarch権限コネクション」を参照してください。

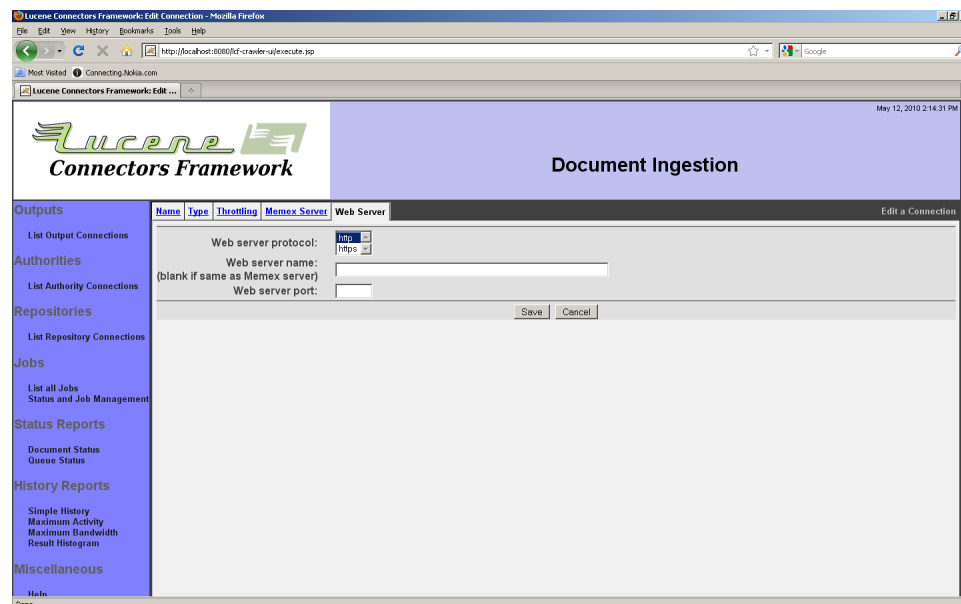
リポジトリコネクション編集ページからMemexコネクションを選択すると次のタブが表示されます:「Memexサーバ」、「Webサーバ」。「Memexサーバ」タブを選択すると次のようなページが表示されます:





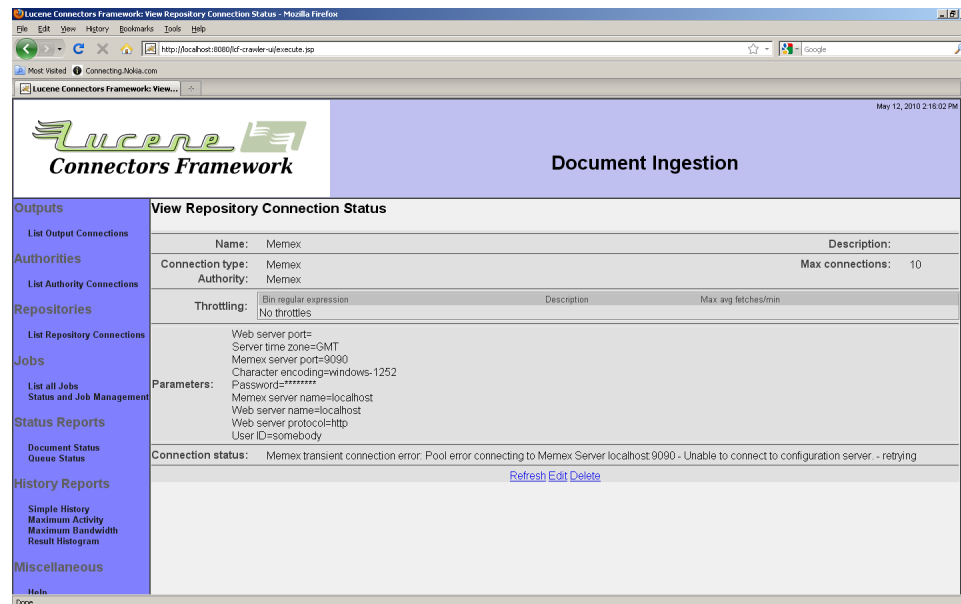
Memexサーバ名、コネクションポート番号、コンテンツを取得できる権限をもつMemexユーザの認証情報を入力して下さい。また、Memexサーバで利用している文字エンコーディングとタイムゾーンも入力してください。

「Webサーバ」タブを選択すると次のようなページが表示されます：



Memexコンテンツ毎に一意的URLを作成できる情報を入力してください。プロトコルを選択して、サーバ名とポート番号を入力してください。

入力した後に「保存」ボタンを押下してください。次のような状態ページが表示されます：



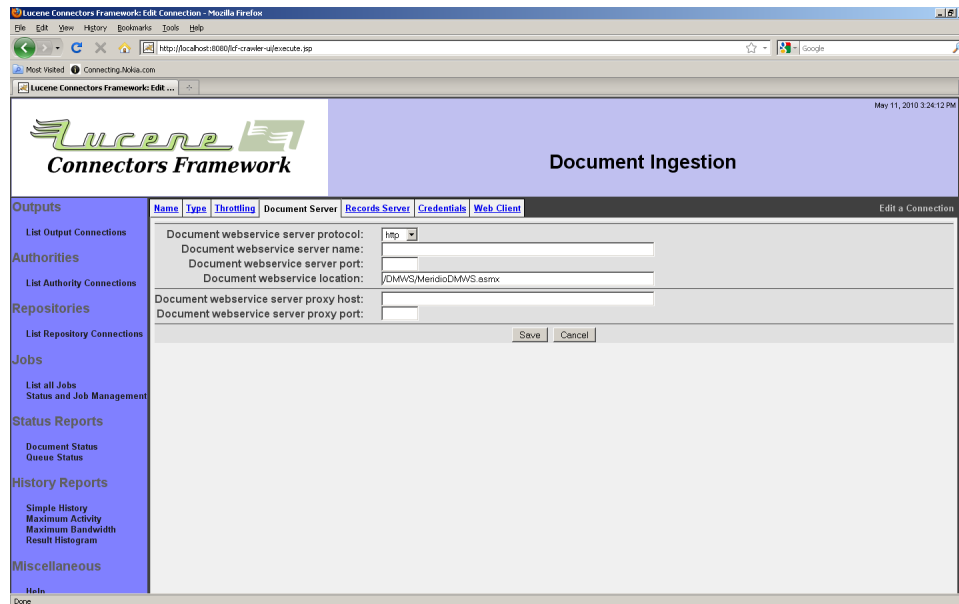
ジョブからMemexコネクションを選択すると次のタブが表示されます：「レコード条件」、「エンティティ」、「セキュリティ」。

### 5.11 Meridioリポジトリコネクション

Autonomy Meridioコネクションは、Meridioサーバのコンテンツから索引を作成します。Meridioのアーキテクチャはサービスを複数のサーバに分散させることを可能にしています（例：ドキュメントサービスを一つのサーバで稼働させ、レコードサービスを別のサーバで稼働させる）。そのため、Meridioコネクションタイプでは、Meridioサーバ毎に設定を行えるようになっています。

Meridioコネクションで処理するコンテンツのセキュリティはMeridio権限を利用します。Meridio権限コネクションについては「Meridio権限コネクション」を参照してください。

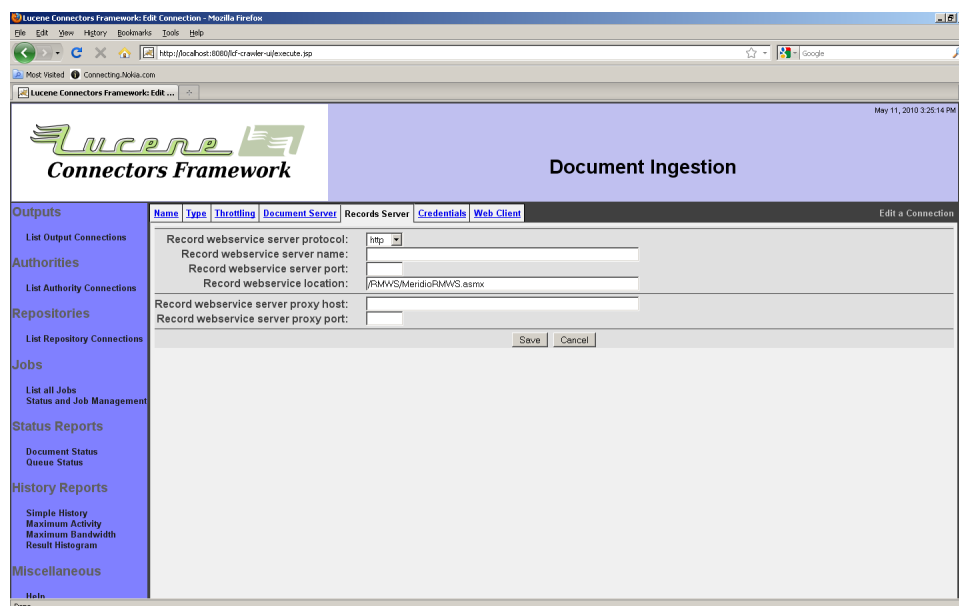
リポジトリコネクションページからMeridioコネクションを選択すると次のタブが表示します：「ドキュメントサーバ」、「レコードサーバ」、「Webクライアント」、「認証」。「ドキュメントサーバ」タブを選択すると次のようなページが表示します：



プロトコルを選択してサーバ名、ポート番号、Meridioドキュメントサーバサービスのアドレスを入力してください。プロキシを利用している場合は、プロキシホストアドレスとポート番号を入力してください。認証プロキシは現リリースでは未対応です。

Meridioシステムの場合は異なるサービス毎にサーバを設けることができますが、一般には複数のサービスが同じサーバで動作しています。ただし、コネクションタイプ設定からは異なるサーバを指定することもできます。

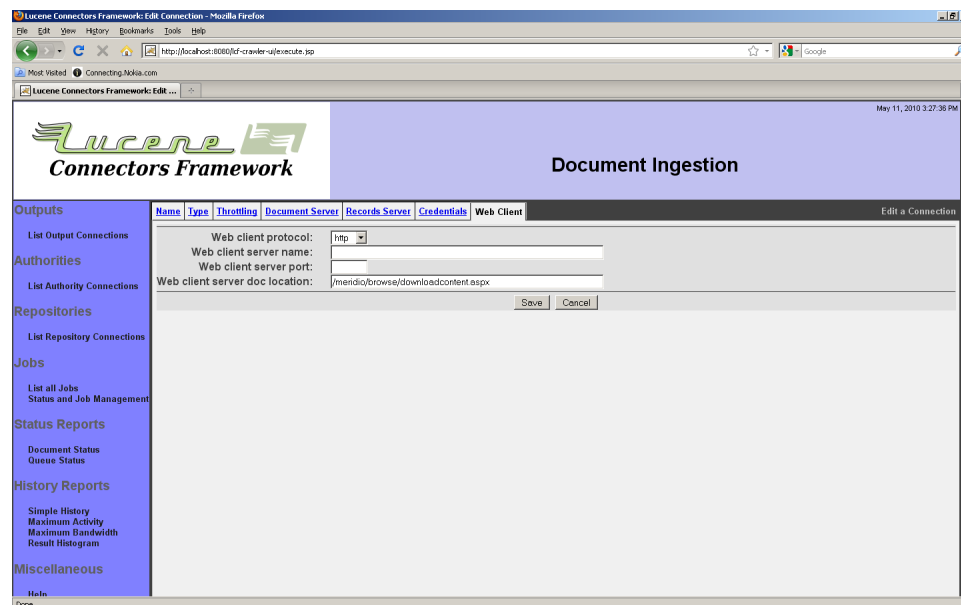
「レコードサーバ」タブを選択すると、次のようなページが表示されます：



プロトコル、サーバ名、ポート番号、Meridioレコードサーバサービスのアドレスを入力してください。プロキシを利用している場合は、プロキシホストとポート番号も入力してください。認証プロキシは現リリースでは未対応です。

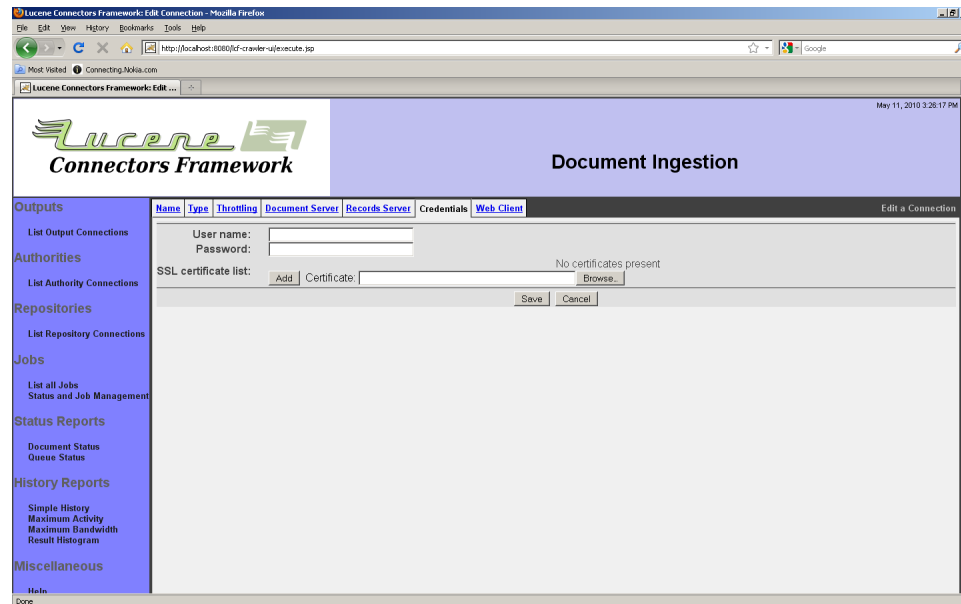
Meridioシステムの場合は異なるサービス毎にサーバを設けることができますが、一般には複数のサービスが同じサーバで動作しています。ただし、コネクションタイプ設定からは異なるサーバを指定することもできます。

「Webクライアント」タブを選択すると次のようなページが表示されます：

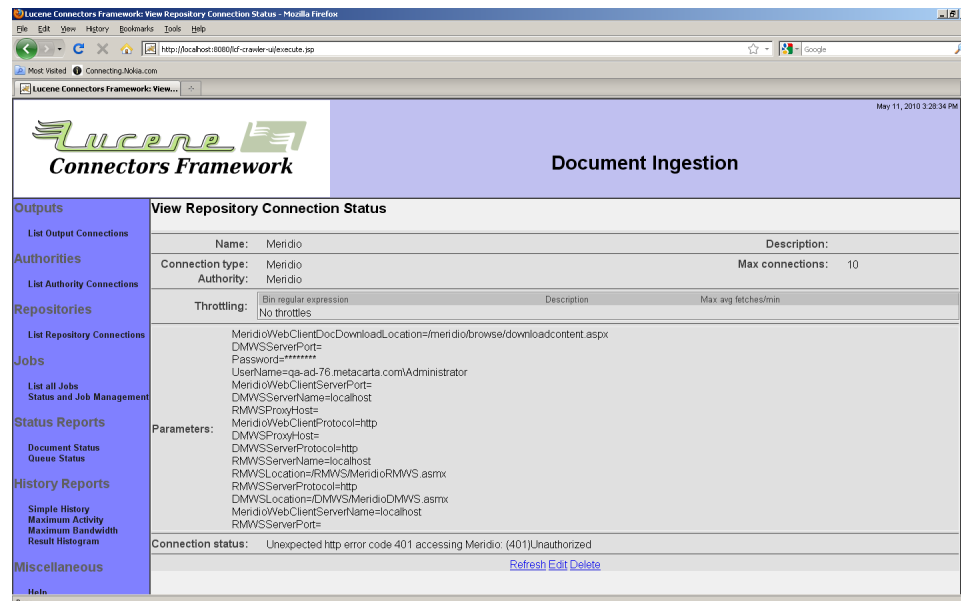


Meridioコネクションwebクライアントタブからは索引を作成したコンテンツ毎にURIを作成します。プロトコルを選択して、サーバ名、ポート番号、Meridio webクライアントサービスのアドレスを入力してください。このサービスからコンテンツを取得しないため、プロキシ情報は不要です。

「認証」タブを選択すると、以下のようなページが表示されます：



Meridioシステム用ユーザの認証情報を入力してください。  
入力したら、「保存」ボタンを押下してください。次のようなページが表示されます：



表示されている画面ではMeridio権限サーバがWindowsドメインに接続できないためエラーになっています。

MeridioはWindows IISの認証機能を利用します。IIS及びWindowsドメインが正しく設定されていない場合は、Meridioも正常に動作しない場合があります。問題が発生した場合

は、Meridio担当技術者に問い合わせてください。また、以下のようなデバッグツールを使うこともできます：

- Windowsセキュリティイベントログ
- ManifoldCFログ(以下の参照)
- パケットキャプチャ(WireSharkなどのツールを利用)

その他のManifoldCFログ情報が必要な場合はソフトウェアを修正する必要があります。

ジョブからMeridioコネクションを選択した場合は次のタブが表示されます：「検索パス」、「コンテンツタイプ」、「分類」、「データタイプ」、「セキュリティ」、「メタデータ」。

## 5.12 Microsoft SharePointリポジトリコネクション

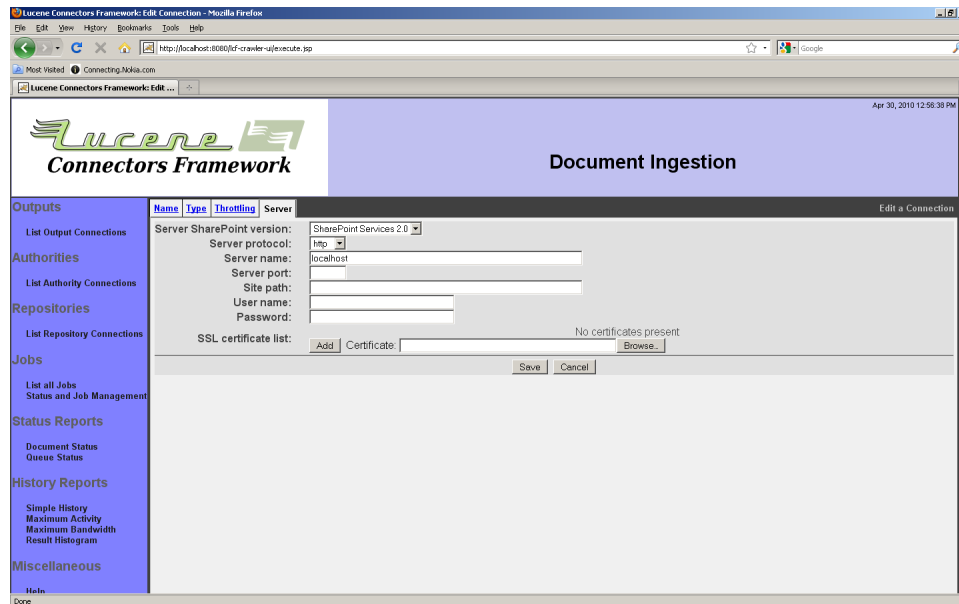
Microsoft SharePointコネクションタイプは、Microsoft SharePointサイトのコンテンツの索引を作成します。SharePointサーバに複数のサイトを構築することができます。SharePointには関連しているサイト(例えばサブサイトの場合)と単独なサイトがあります。

SharePointコネクションタイプは1つのSharePointリポジトリコネクションからルートサイトの明示的なサブサイトを介してすべてのSharePointサイトをアクセスできるように設計されています。大きなSharePointサイトの場合はこのような1つのコネクションからすべてのSharePointサイトをアクセスできるようにする要求が高いです。ただし現バージョンのManifoldCFでは未対応です。

SharePointはサイト、サブサイト、ライブラリ、ファイルをアドレスにweb URIを利用します。そのため、SharePointコネクションタイプを調べる一番よい方法はwebブラウザからクロールしたいサイトのルートからページを開いていき、URLを記録することです。

多くの場合は、SharePointコネクションで処理されるコンテンツのセキュリティはActive Directoryで管理されています。Active Directory権限を作成していない場合は、「Active Directory権限コネクション」を参照してください。

リポジトリコネクション編集からSharePointコネクションを選択すると「サーバ」タブが表示します。「サーバ」タブを選択すると次のようなページが表示されます：



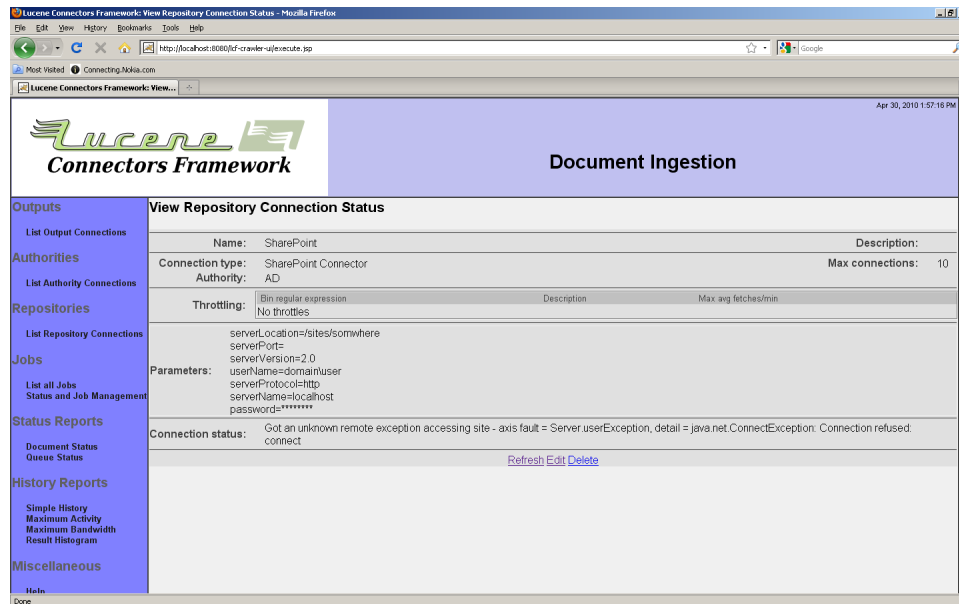
プロダクトリストからSharePointサーバのバージョンを選択してください。間違ったバージョンを選択すると、索引ができなかったり権限情報が正しく取得されない場合があります。

サーバのプロトコルを選択してSharePointサイトを調べて記録したURIからサーバ名とポート番号を入力してください。項目「サイトパス」にはルートサイトURIのサーバアドレスとポート番号以降の最後の「.aspx」ファイル以外の文字列を入力してください。例えばSharePoint URIが「http://myserver:81/sites/somewhere/index.aspx」の場合は、サイトパスは「/sites/somewhere」です。

SharePointの認証を使ってルートサイトにログインします。SharePointコネクションタイプのユーザ名は必ず「domain¥user」形式で記入してください。

SharePointサーバがSSLを利用している場合は、SharePointサーバのSSLサーバ認証サーバ証明書または認証局からの証明書を設定してください。参照から証明書を選択して、「追加」ボタンを押下してください。

「保存」ボタンを押下すると次のような接続設定概要ページが表示されます：



画面例ではSharePointコネクションはSharePointインスタンスを参照できないためエラーメッセージが表示されています。

SharePointは認証にWindows IISを利用します。SharePointが稼動しているIIS及びWindowsドメインでの問題のためSharePointコネクションが正常に動作しない場合もありますので注意してください。問題が発生した場合は次のようなデバッグツールを使うことができます：

- Windowsセキュリティイベントログ
- ManifoldCFログ(以下の参照)
- パケットキャプチャ(WireSharkなどのツールを利用)

標準以外のログ情報が必要な場合はソフトウェアを修正する必要があります。

SharePointリポジトリコネクションを選択した場合は、次のようなタブが表示されます：「パス」、「セキュリティ」、「メタデータ」。

「パス」タブからはジョブに含む／除外するSharePointコンテンツを指定するルール一覧を作成ことができます。SharePointコネクションタイプがサブサイト、ライブラリ、ファイルを見つけた場合は、このルール一覧を参照して含むか除外するかを判断します。最初に一致したルールが適用されます。

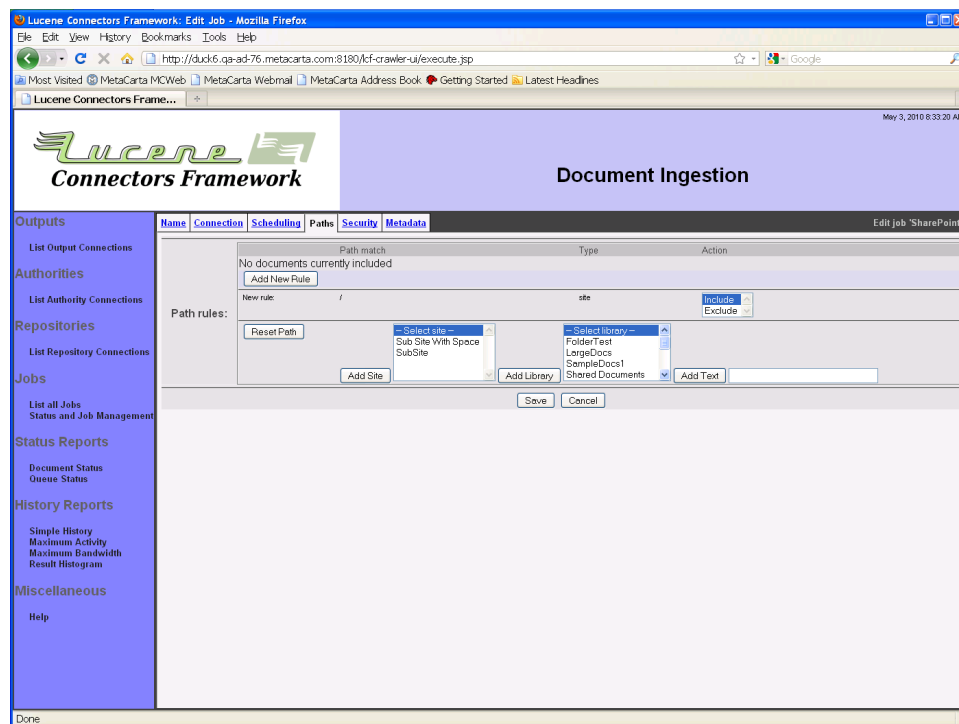
各ルールはパス、ルールタイプ、処理から構成されます。処理とは「含む」か「除外」するかです。ルールタイプはコネクションにどのタイプのSharePointエンティティかを識別します。例えば、「ファイル」ルールはSharePointパスがファイルの場合のみ一致して、サイト及びライブラリには一致しません。パスは文字列です。ワイルドカード文字「\*」と「?」を使うこともでき



ます。「\*」は0以上の任意の文字と一致します。「?」は任意の1文字と一致します。その他の文字は記述通りに一致する必要があります。

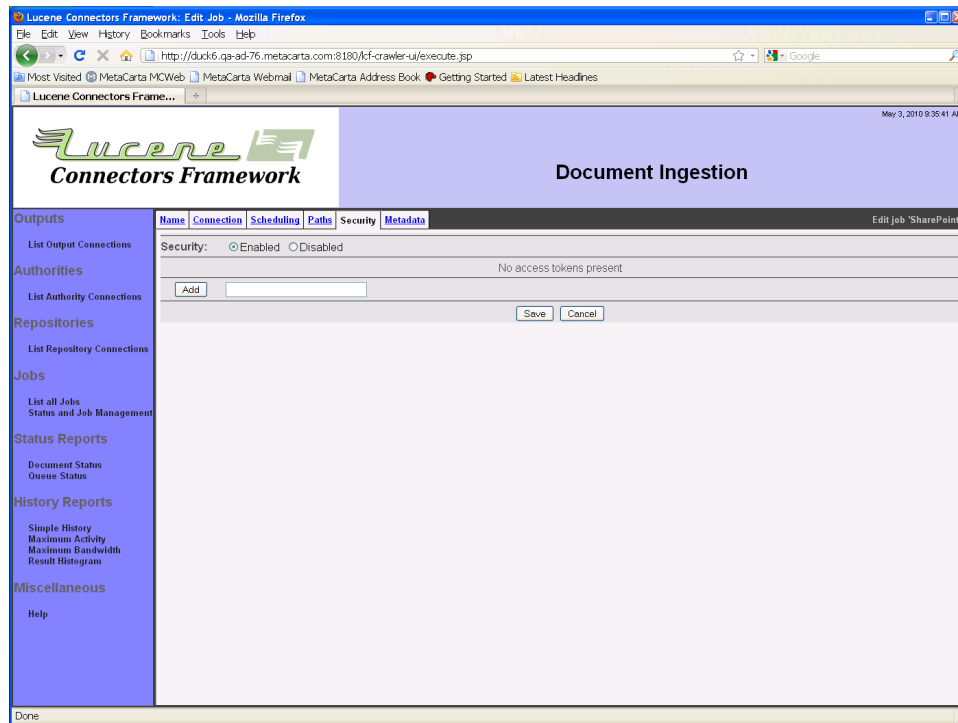
暗黙的に一致するルールも定義することができます。「含む」の「ファイル」を選択した場合は、サイトとライブラリも暗黙的に「含む」になります。例えば、「/MySite/MyLibrary/MyFile」を「含む」ルールを定義した場合は、暗黙的に「/MySite」サイトを含むサイトルールと「/MySite/MyLibrary」を含むライブラリルールも定義されます。同じようにライブラリを含むルールを定義した場合はサイトを含むルールも定義されます。これらの暗黙ルールは「含む」ルールのみで定義されます。除外ルールには暗黙ルールはありません。

「パス」タブからルールを作成して一覧に追加／挿入することができます。「パス」タブを選択すると次のようなページが表示されます。コネクションが正常に動作していない場合は、プルダウンに表示する項目が少ない場合もあります。



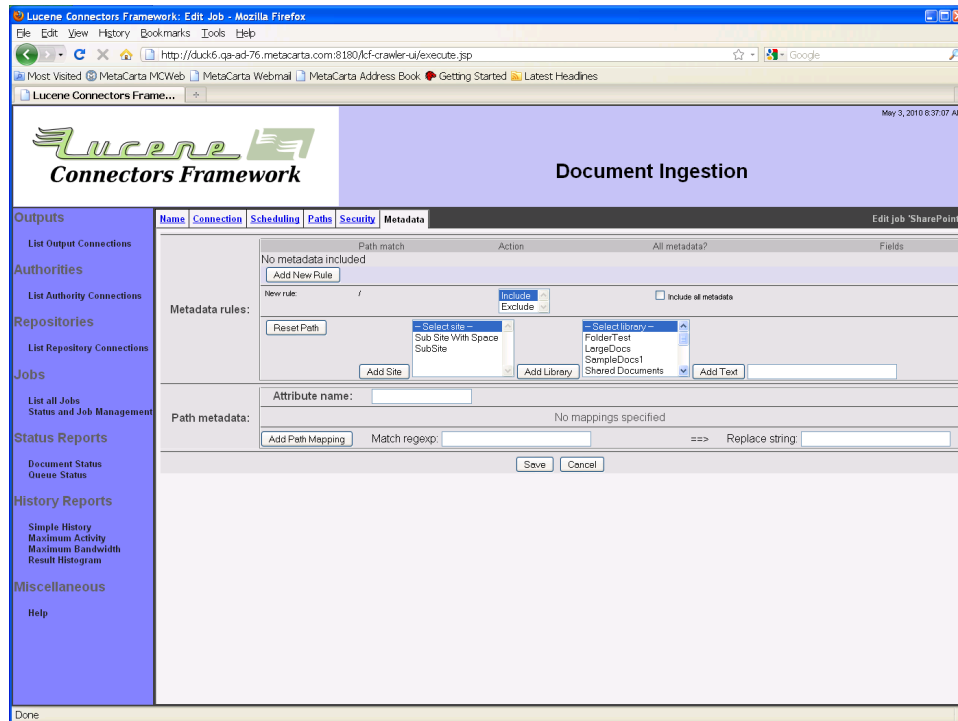
ルールを定義するには、まず一致するパスを指定します。パスを選択または入力して「サイトの追加」ボタン、「ライブラリの追加」ボタン、「テキストの追加」ボタンの一つを押下してください。完全のパスを指定するまで、追加を繰り返してください。SharePointコネクションがパスのエンティティを判断できない場合は、SharePointエンティティを手動で選択してください。次に「含む」または「除外」用のルールかを選択してください。最後に「新しいルールを追加」ボタンを押下してルールをルール一覧に追加してください。

「セキュリティ」タブからSharePointセキュリティをコンテンツに適用するか指定することができます。ジョブが対象とするコンテンツにアクセストークンを付けることもできます。「セキュリティ」タブを選択すると次のようなページが表示されます：



SharePointセキュリティ有効／無効のラジオボタンを選択してください。セキュリティを無効にした場合は、アクセストークンを設定することができます。アクセストークンを入力して「追加」ボタンを押下してください。アクセストークンはSharePointコネクションが利用できる形式にしてください。

「メタデータ」タブからは各コンテンツに含むメタデータを指定することができます。「メタデータ」タブの内容は「パス」タブと類似しています。「メタデータ」タブを選択すると次のようなページが表示されます：



「パス」タブとの違いは、個々のサイト、ライブラリ、コンテンツを含む／除外するルールを定義する代わりにコンテンツを含む／除外するメタデータを定義します。メタデータはファイルのみに提供可能なため、サイト及びライブラリ用のメタデータパスルールはありません。

除外ルールがファイルパスに一致した場合は、そのファイルからのすべてのメタデータが除外されます。除外ルールで個々の項目を除外することはできません。

ルールを定義するには、まず一致するパスを指定します。パスを選択または入力して「サイトの追加」ボタン、「ライブラリの追加」ボタン、「テキストの追加」ボタンの一つを押下してください。完全のパスを指定するまで、追加を繰り返してください。SharePointコネクションがパスのエンティティを判断できない場合は、SharePointエンティティを手動で選択してください。次に「含む」または「除外」用のルールかを選択してください。最後に「新しいルールを追加」ボタンを押下してルールをルール一覧に追加してください。

「パスメタデータ」タブからはコンテンツ毎のパス情報を索引にメタデータとして送るように指定することができます。送るようにする場合は、項目「パス属性名」にメタデータ属性名を入力して、ルールをルール一覧に追加してください。各ルールに一致する正規表現の式で構成されます。変換元と値は括弧（「(」と「)」）で囲みます。括弧に囲まれた部分を「グループ」と言います。置き換え文字列は、固定文字と置き換えグループから構成されます。例えば、「\$(1)」は最初に一致したグループを示し、「\$(1l)」は最初に一致した小文字のグループを示します。同じように「\$(1u)」は大文字にマップしたグループを示します。

例えば、ルールの一一致条件が`*/(.*)/(.*)/*`で置き換え文字列`$ (1) $ (2)/`とした場合、パス`Project/Folder_1/Folder_2/Filename`は`Folder_1 Folder_2`に変換されます。1つ以上のルールが存在する場合は、上から実行され、上のルールの結果は下のルールで変更されます。

### 5.13 CMISリポジトリコネクション

CMISリポジトリコネクションタイプは、CMIS準拠リポジトリのコンテンツの索引を作成します。

デフォルト設定では、各CMISコネクションは一つのCMISリポジトリを処理します。複数のCMISリポジトリがある場合は、CMISリポジトリ毎にCMISコネクションを作成する必要があります。

リポジトリコネクション編集ページからCMISコネクションを選択すると次のような項目を設定することができます：

CMISバインディングプロトコル (AtomPubまたはWeb Service) を選択して、ユーザ名、パスワード、CMISドキュメントサーバサービスのエンドポイントを入力してください。

エンドポイントはCMISサービスのHTTPプロトコル、ホスト名、ポート番号、コンテキストパスから構成されます：

`http://ホスト名:ポート番号/CMISコンテキストパス`

公開されているCMISリポジトリを一つ選択するためにリポジトリIDを入力することもできます。もしnullの場合は、CMISコネクタはCMISサーバが公開している最初のCMISリポジトリを利用します。

CMISシステムは特定のバインディングプロトコルは独自のコンテキストパスがあります。即ち、エンドポイントは異なります：

OpenCMISが提供する実際のInMemoryサーバが公開するAtomPubバイディングのエンドポイントは次の通りです:

`http://localhost:8080/chemistry-opencmis-server-inmemory-war-0.5.0-SNAPSHOT/atom`

Web Serviceは別のエンドポイントで公開されます:

`http://localhost:8080/chemistry-opencmis-server-inmemory-war-0.5.0-SNAPSHOT/services/RepositoryService`

「保存」ボタンを押下すると、次のようなコネクション概要ページが表示されます:

The screenshot shows the 'View Repository Connection Status' page in the Apache ManifoldCF interface. The page title is 'Document Ingestion'. The main content area displays the following information:

- Name:** CMIS Ingestion
- Description:** Ingestion from a CMIS repository
- Connection type:** CMIS
- Authority:** None (global authority)
- Max connections:** 10
- Throttling:** No throttles
- Parameters:**
  - username=dummyuser
  - binding=atom
  - password=\*\*\*\*\*
  - endpoint=http://localhost:8080/chemistry-opencmis-server-inmemory-war-0.5.0-SNAPSHOT/atom
- Connection status:** Connection working

At the bottom of the main content area, there are links for 'Refresh', 'Edit', and 'Delete'. The left sidebar contains the following navigation links:

- Outputs
  - List Output Connections
- Authorities
  - List Authority Connections
- Repositories
  - List Repository Connections
- Jobs
  - List all Jobs
  - Status and Job Management
- Status Reports
  - Document Status
  - Queue Status
- History Reports
  - Simple History
  - Maximum Activity
  - Maximum Bandwidth
  - Result Histogram
- Miscellaneous
  - Help

ジョブでCMISリポジトリコネクションを選択すると「CMISクエリー」タブが表示されます。「CMISクエリー」タブを選択すると次のようなページが表示されます:

The screenshot shows the 'CMIS Query' page in the Apache ManifoldCF interface. The page title is 'Document Ingestion'. The main content area displays the following information:

- CMIS Query:** SELECT \* FROM cmis:folder WHERE cmis:name='testdata'
- Buttons:** Save, Cancel

The left sidebar is the same as the previous screenshot.

「CMISクエリー」タブからはコンテンツを取得するCMISクエリー言語のクエリー文を設定することができます。

CMISコネクタはコンテンツの読み込み中にフォルダノード (baseTypeがcmis:folderのノード) を見つけた場合は、フォルダ内のコンテンツも読み込みます。フォルダではない場合 (baseTypeがcmis:document) は、コンテンツを読み込んで処理します。

入力した後は「保存」ボタンを押下してください。次のように設定概要が表示されます：

The screenshot shows the Apache ManifoldCF web interface. The top header is 'Document Ingestion'. The left sidebar contains navigation links: Outputs, Authorities, Repositories, Jobs, Status Reports, History Reports, Miscellaneous, and Help. The main content area is titled 'View a Job' and displays configuration details for a 'CMIS Ingestion Job'.

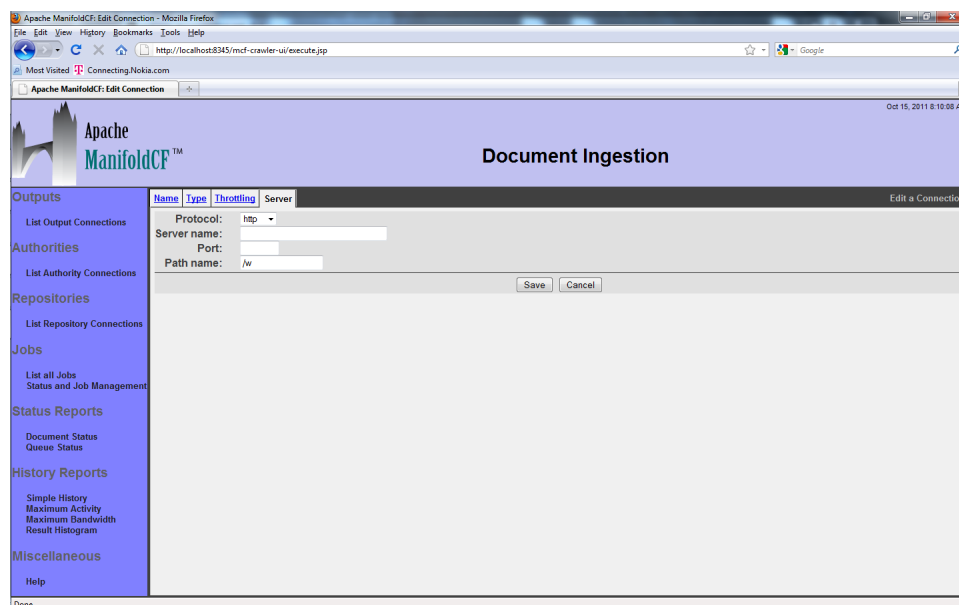
Name: CMIS Ingestion Job			
Output connection: Null Output		Repository connection: CMIS Ingestion	
Priority: 5		Start method: Don't automatically start	
Schedule type: Scan every document once		Minimum recrawl interval: Not applicable	
Expiration interval: Not applicable		Reseed interval: Not applicable	
No scheduled run times			
Maximum hop count for link type 'child': Unlimited			
Hop count mode: Delete unreachable documents			
CMIS Query: SELECT * FROM cmis:folder WHERE cmis:name='testdata'			
<a href="#">Edit</a> <a href="#">Delete</a> <a href="#">Copy</a>			

## 5.14 Wikiリポジトリコネクション

Wikiリポジトリコネクションタイプは、Wiki及びMediaWikiサイトのコンテンツから索引を作成します。WikiリポジトリコネクションタイプはWiki APIを介してコンテンツを取得します。一般公開されている内容のみから索引を作成するため、認証設定はありません。

このコネクションタイプでは、ジョブごとに手入力されたアクセストークン以外のいかなる種類の文書セキュリティもサポートされていません。

リポジトリ編集ページで、Wikiコネクションを選択すると、「サーバ」タブが表示されます。「サーバ」タブを選択すると次のようなページが表示されます：



プロトコルは「プロトコル」ドロップダウンリストから選択してください。現バージョンは「http」プロトコルのみに対応しています。サーバ名を項目「サーバ名」に、ポート番号を項目「ポート」に入力してください。最後に、WikiのURLのパス部分を項目「パス名」に入力してください。URLのパス部分の先頭は文字「/」にしてください。

ジョブのリポジトリコネクションにはWikiタイプ固有のタブは現バージョンにはありません。