ManifoldCF- エンドユーザマニュアル

Table of contents

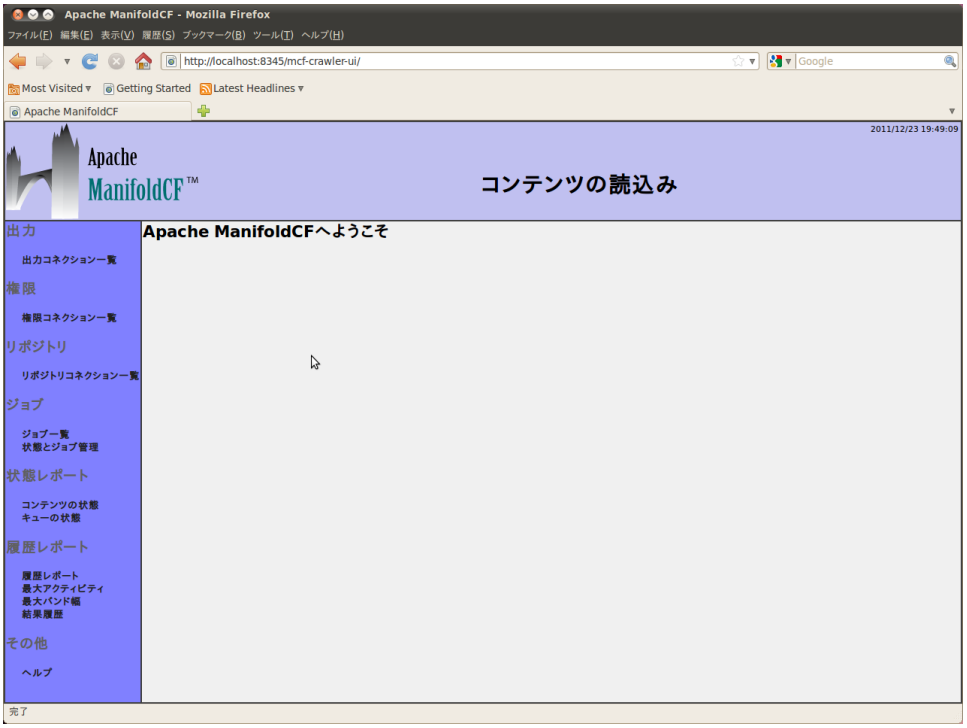
1 概要

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付いて説明します。デフォルト設定では、Webブラウザで次のURLを開きます：http://my-server-name:8345/mcf-crawler-ui。デフォルト設定ではない場合は、異なるURLの場合があります。システム管理者に問い合わせてください。

ManifoldCFのユーザインタフェースはFirefox及び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。正しく動作しない場合は、システム管理者に連絡してください。

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：

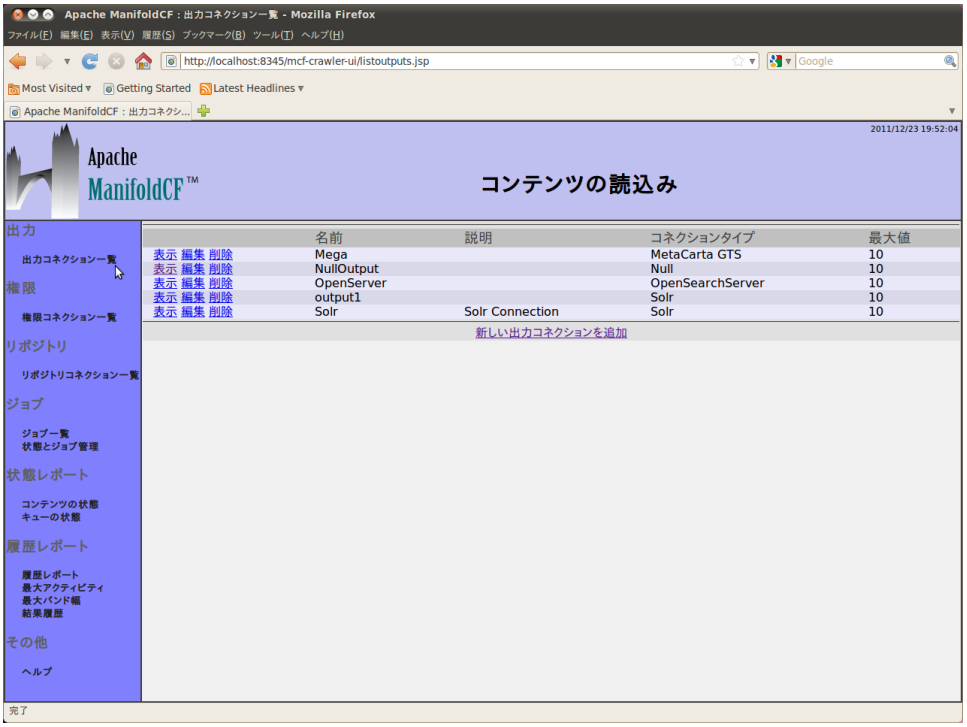


左側にメニューが表示されます。ページを開くと、右には挨拶メッセージが表示されます。メニューから項目を選択すると、右側に表示される内容が変わります。設定を行う前に、下の説明に目を通してManifoldCFの概要を理解することをお勧めします。

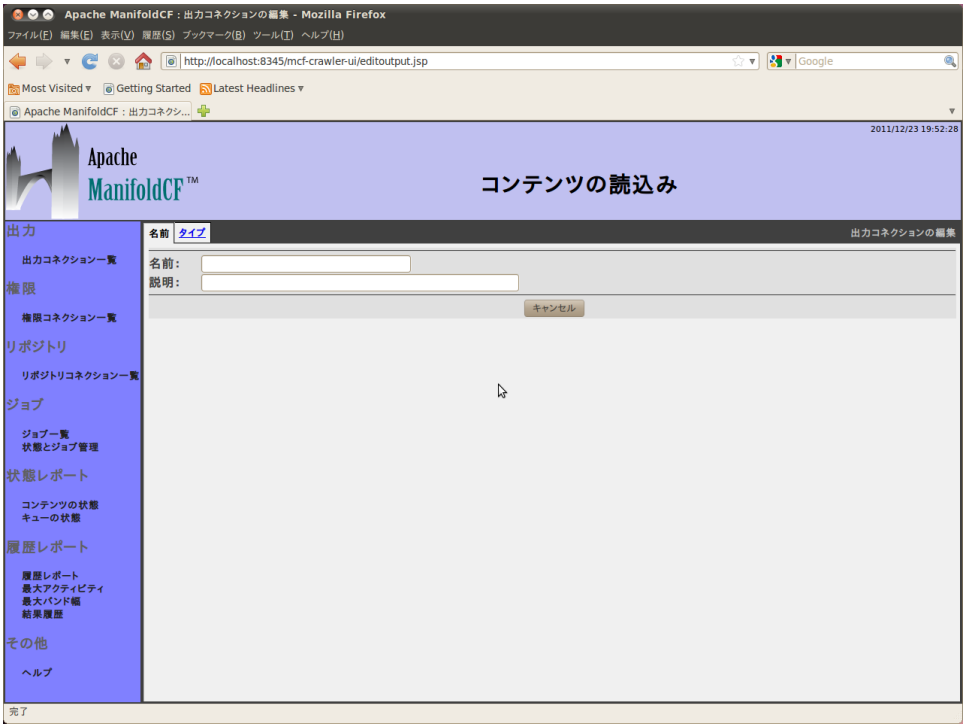
1.1 出力コネクタの定義

左側のメニューには読み込んだコンテンツを出力する出力先コネクタの一覧があります。多くの場合は、検索エンジンに出力されます。

すべてのジョブには出力コネクションを指定する必要があります。出力コネクションを指定するには、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。以下のようなページが表示されます：

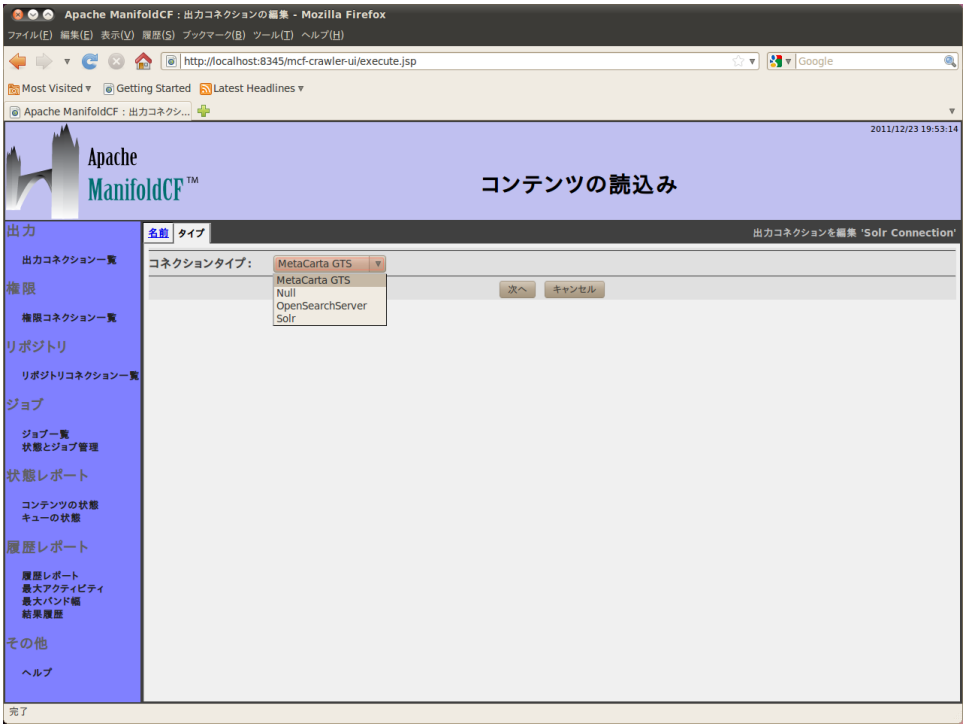


始めて利用される場合は、出力コネクションは定義されていないかもしれません。出力コネクションが定義されている場合は、一覧表示されます。出力コネクションの左に表示される「表示」、「編集」、「削除」リンクを選択して設定内容を表示、編集したり削除することができます。新しい出力コネクションを定義する場合は、一覧の下の「新しい出力コネクションを追加」リンクを選択してください。選択すると以下のようなページが表示されます：



上に表示されるタブは出力コネクションの要素の纏まりです。コネクションタイプによって表示されるタブは異なります。

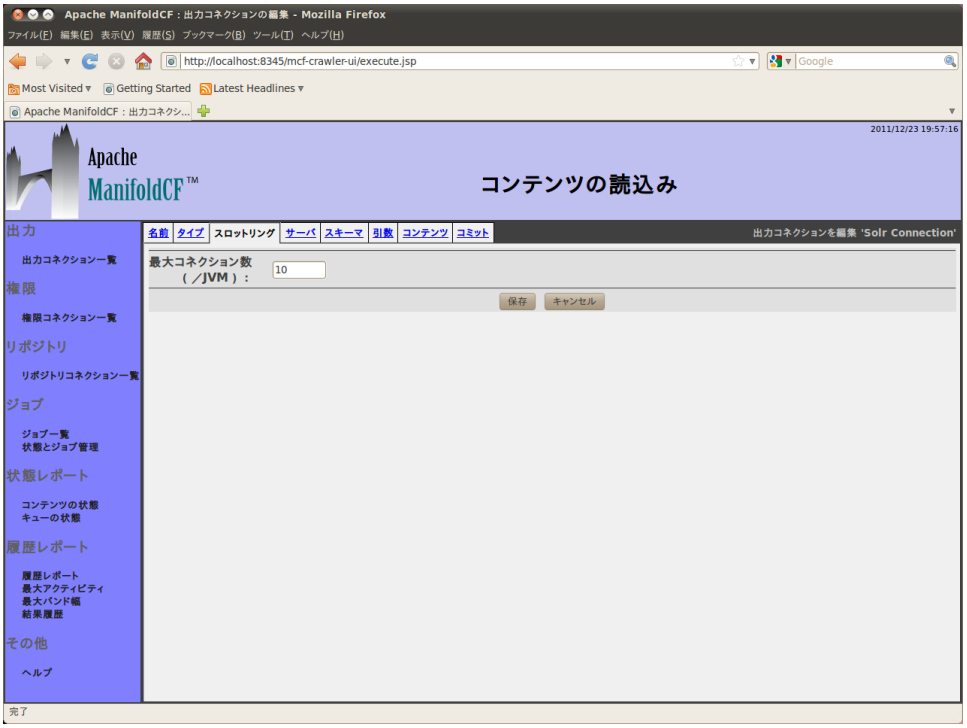
先ず、コネクションに付ける名前と説明を入力してください。出力コネクション名はコネクションを識別するために使われるため、一意性である必要があります。また、一旦設定すると変更することができませんので注意してください。名前は32文字以内、説明は255文字以内に設定してください。入力した後に、「タイプ」タブを選択してください。選択すると以下のようなページが表示されます：



コネクションタイプ・プルダウンリストを選択すると、出力コネクション一覧が表示されます。一覧に表示される出力コネクタ及び名前は、Apache ManifoldCFをンストール／セットアップしたシステムインテグレータによって異なる場合があります。表示されるタブは、選択されたコネクションタイプによって変わります。以降の節で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プルダウンリストから選択して、「次へ」ボタンを選択してください。選択された出力コネクションの定義に必要な要素の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。設定内容を破棄する場合は、「キャンセル」ボタン又は左に表示されるメニュー項目を選択してください。

すべての出力コネクションタイプには「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：



このページには一つの項目のみがあります: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。システムの負荷を調整したり、ライセンス制約によるコネクション数の制限を行うことができます。値を大きくすると、スループットが上がります。コネクションタイプすべてのデフォルト値は10ですが、出力コネクション・タイプによってはこの値は最適ではない場合もあります。詳細については、以降の各種の出力コネクション・タイプの説明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ページが表示されます。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「正常」と表示されます。設定に間違いがある場合は、エラー内容が表示されます。エラーメッセージが表示された場合は、設定内容を修正してください。

1.2 権限コネクションの定義

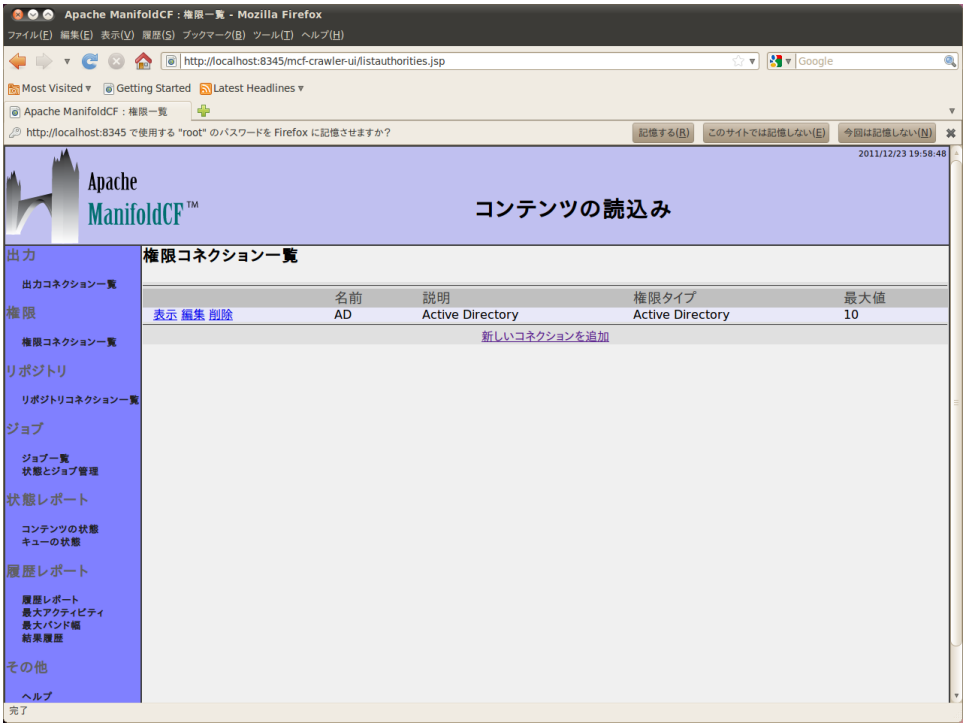
左側メニューから「権限一覧」を選択すると、権限コネクション一覧が表示されます。権限コネクションとは、特定のセキュリティ環境をもつシステムへ接続するためのコネクションです。例えば、アクティブディレクトリで管理されているファイルを参照する場合は、アクティブディレクトリ権限コネクションを定義します。

一般公開されているコンテンツのみをクロールする場合は、権限コネクションを定義する必要はありません。例えばインターネット上で認証なしでだれでも閲覧できるwebページ, RSS, Wikiをクロールする場合は不要です。反対に、社内で利用されている多くのコンテンツ

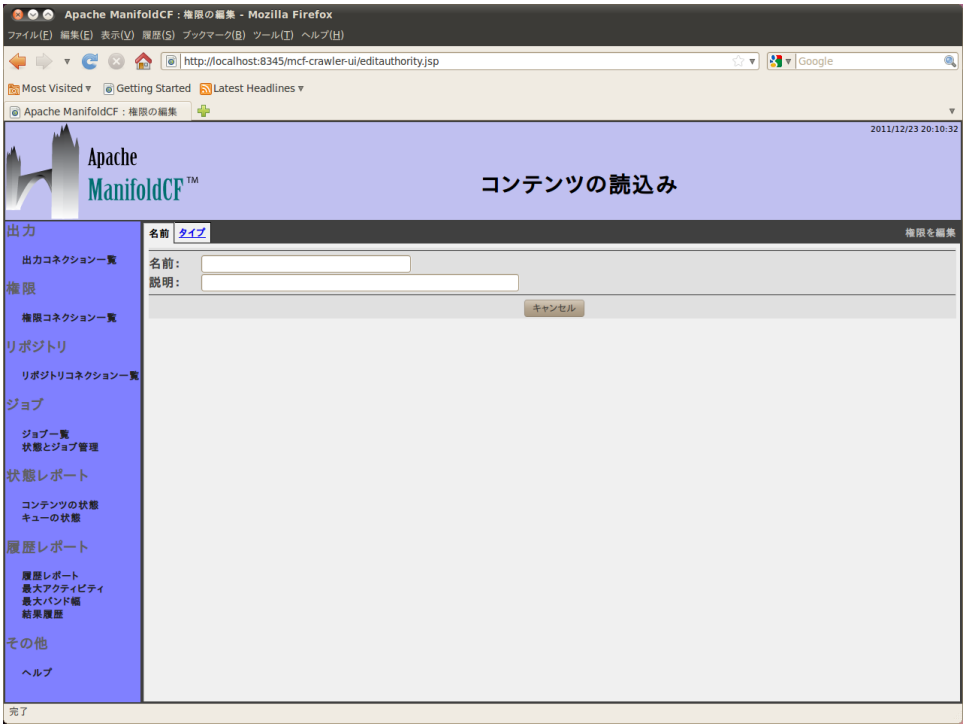
を検索する場合はユーザ認証が必要な場合が多いため、権限コネクションを定義する必要があります。

リポジトリコネクションを定義する前に権限コネクションを定義してください。後でリポジトリコネクションと権限コネクションの関係を変更することも可能ですが、変更した場合はコンテンツを再クロールされる必要があるかもしれません。

権限コネクションを定義するには、右側メニューから「権限一覧」を選択してください。選択すると次のようなページが表示されます：

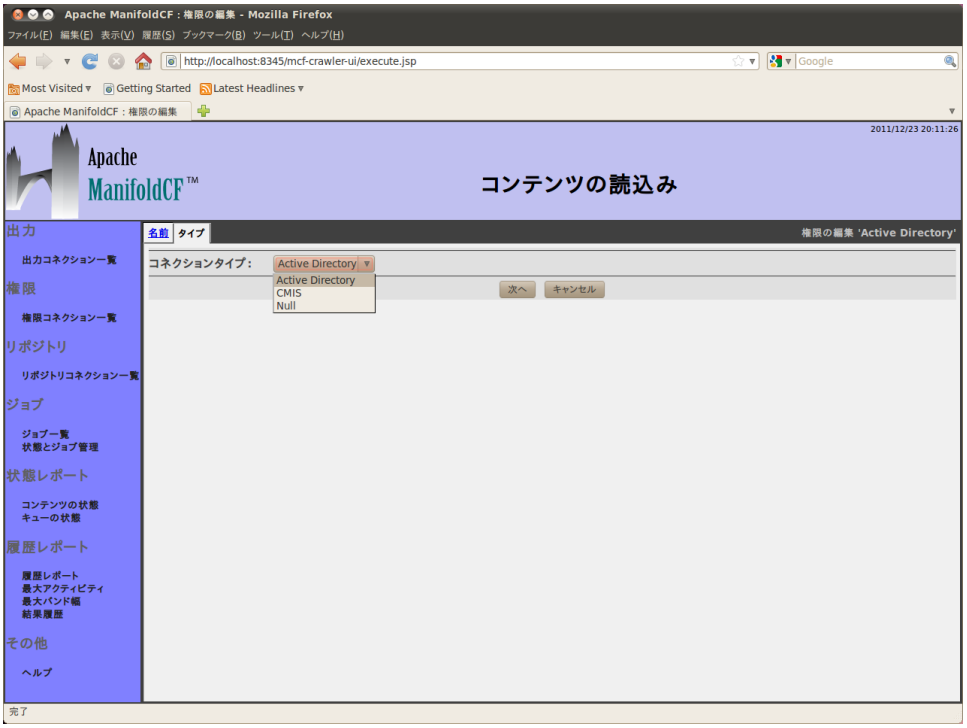


始めて利用される場合は、権限コネクションは定義されていないかもしれません。権限コネクションが定義されている場合は、一覧表示されます。権限コネクションの左に表示される「表示」、「編集」、「削除」リンクを選択して設定内容を表示、編集したり削除することができます。新しい権限コネクションを定義する場合は、一覧の下の「新しい権限コネクションを追加」リンクを選択してください。選択すると以下のようなページが表示されます：



上に表示されるタブは権限コネクションの要素の纏まりです。コネクションタイプによって表示されるタブは異なります。

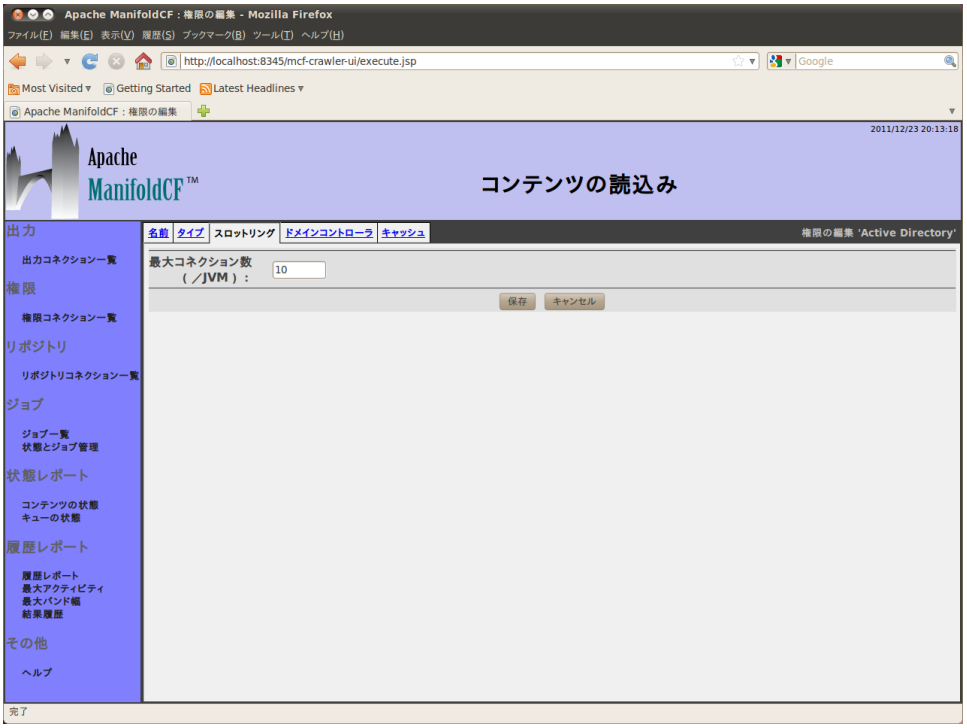
先ず、コネクションに付ける名前と説明を入力してください。権限コネクション名はコネクションを識別するために使われるため、一意性である必要があります。また、一旦設定すると変更することができませんので注意してください。名前は32文字以内、説明は255文字以内に設定してください。入力した後に、「タイプ」タブを選択してください。選択すると以下のようなページが表示されます：



コネクションタイプ・プルダウンリストを選択すると、権限コネクション一覧が表示されます。一覧に表示される権限コネクタ及び名前は、Apache ManifoldCFをンストール／セットアップしたシステムインテグレータによって異なる場合があります。表示されるタブは、選択されたコネクションタイプによって変わります。以降の節でタブの設定内容を説明します。

権限コネクションタイプをプルダウンリストから選択して、「次へ」ボタンを選択してください。選択された権限コネクションの定義に必要な要素の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。設定内容を破棄する場合は、「キャンセル」ボタン又は左に表示されるメニュー項目を選択してください。

すべての出力コネクションタイプには「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：



このページには一つの項目のみがあります：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。システムの負荷を調整したり、ライセンス制約によるコネクション数の制限を行うことができます。値を大きくすると、スループットが上がります。コネクションタイプすべてのデフォルト値は10ですが、出力コネクション・タイプによってはこの値は最適ではない場合もあります。詳細については、以降の各種の出力コネクション・タイプの説明を参照してください。

コネクションタイプのタブの詳細に付いては、権限コネクション・タイプによって表示されるタブの説明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ページが表示されます。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「正常」と表示されます。設定に間違いがある場合は、エラー内容が表示されます。エラーメッセージが表示された場合は、設定内容を修正してください。

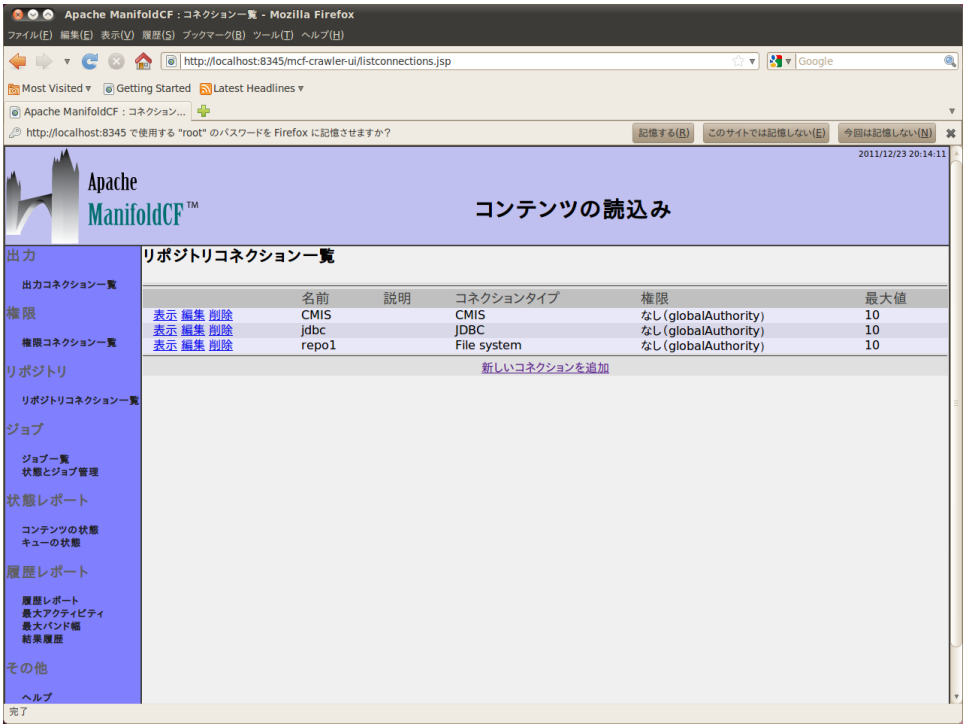
1.3 リポジトリコネクションの定義

左側メニューから「リポジトリ一覧」を選択すると、リポジトリコネクション一覧が表示されます。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へのコネクションです。 特定のセキュリティ環境をもつシステムへ接続するためのコネクションです。例えば、アクティブディレクトリで管理されているファイルを参照する場合は、アクティブディレクトリ権限コネクションを定義します。

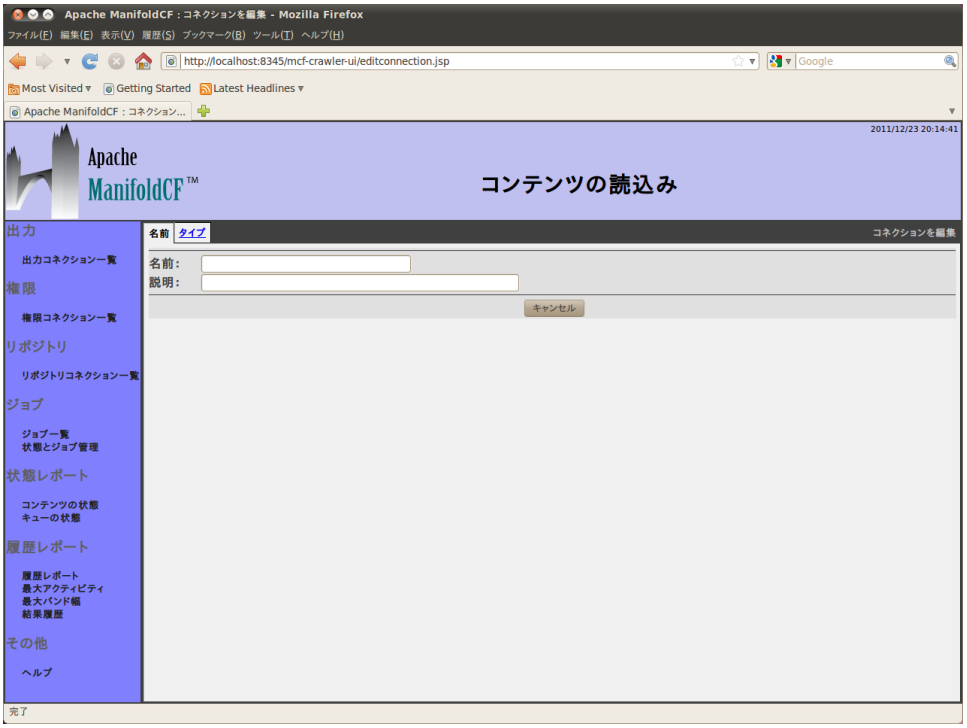
すべてのジョブにはリポジトリコネクションを指定する必要があります。ジョブは指定されたリポジトリコネクションからコンテンツを読み込みます。コンテンツから索引を作成するジョブを定義する前に、リポジトリコネクションを作成してください。

リポジトリコネクションに権限コネクションを指定することもできます。権限コネクションは、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指定します。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コネクションを変更することもできますが、この場合は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を再読み込みして索引を再構成する必要があります。そのため、リポジトリコネクションを定義する前に権限コネクションを定義することを推奨します。スロットリング

リポジトリコネクションを定義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。次のようなページが表示されます：



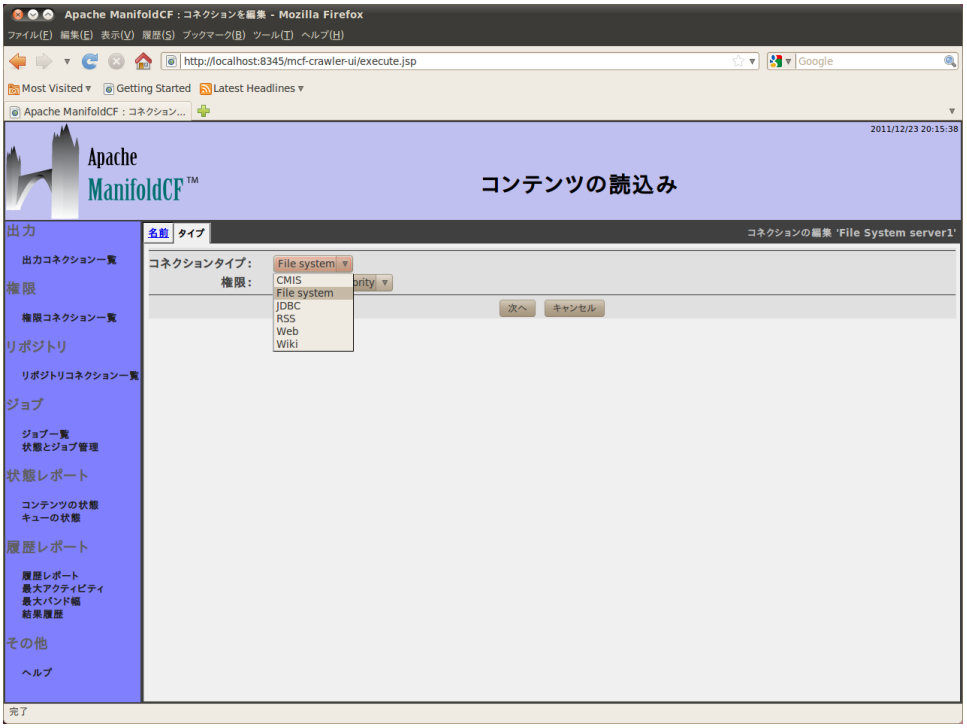
始めて利用される場合は、リポジトリコネクションは定義されていないかもしれません。リポジトリコネクションが定義されている場合は、一覧表示されます。リポジトリコネクションの左に表示される「表示」、「編集」、「削除」リンクを選択して設定内容を表示、編集したり削除することができます。リポジトリしい出力コネクションを定義する場合は、一覧の下の「新しいリポジトリコネクションを追加」リンクを選択してください。選択すると以下のようなページが表示されます：



上に表示されるタブはリポジトリコネクションの要素の纏まりです。コネクションタイプによって表示されるタブは異なります。

先ず、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コネクションを識別するために使われるため、一意性である必要があります。また、一旦設定すると変更することが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文字以内に設定してください。入力した後に、「タイプ」タブ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:

コネクションタイプのタブの詳細に付いては、権限コネクション・タイプによって表示されるタブの説明を参照してください。

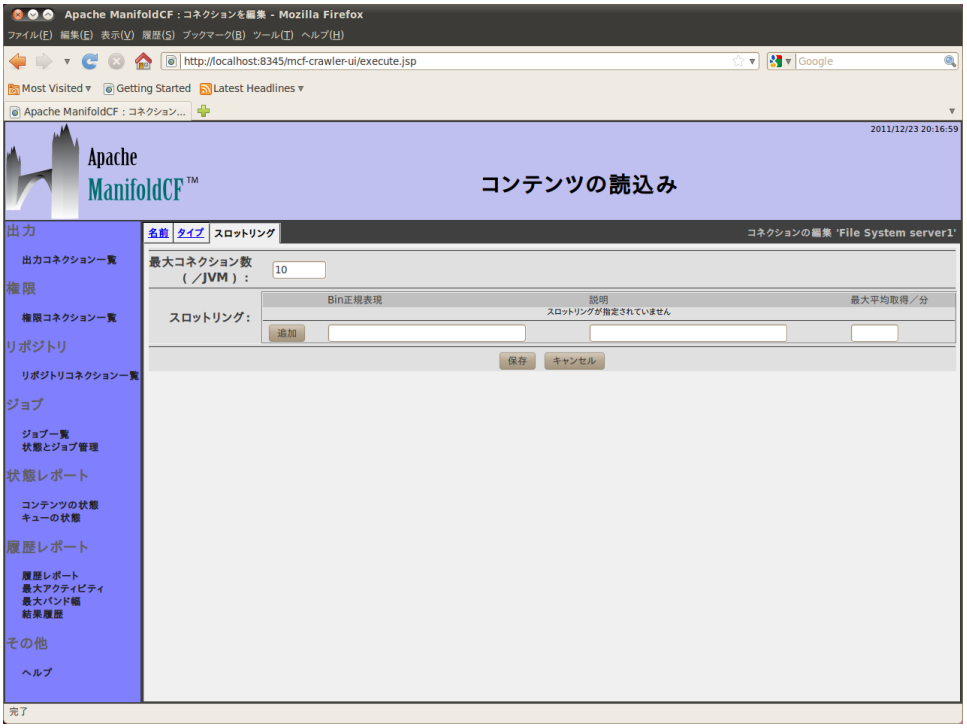


コネクションタイプ・プルダウンリストを選択すると、リポジトリコネクション一覧が表示されます。一覧に表示されるリポジトリコネクタ及び名前は、Apache ManifoldCFをンストール／セットアップしたシステムインテグレータによって異なる場合があります。表示されるタブは、選択されたコネクションタイプによって変わります。以降の節でタブの設定内容を説明します。

リポジトリから読み込むコンテンツの権限情報を指定することもできます。権限コネクションはリポジトリコネクションに依存している場合もありますので注意してください。詳細に付いては利用されるリポジトリコネクション及び権限コネクションの説明を参照してください。

リポジトリコネクションタイプと権限コネクションを選択した後は「次へ」ボタンを選択してください。選択されたリポジトリコネクションの定義に必要な要素の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。設定内容を破棄する場合は、「キャンセル」ボタン又は左に表示されるメニュー項目を選択してください。

すべての出力コネクションタイプには「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：



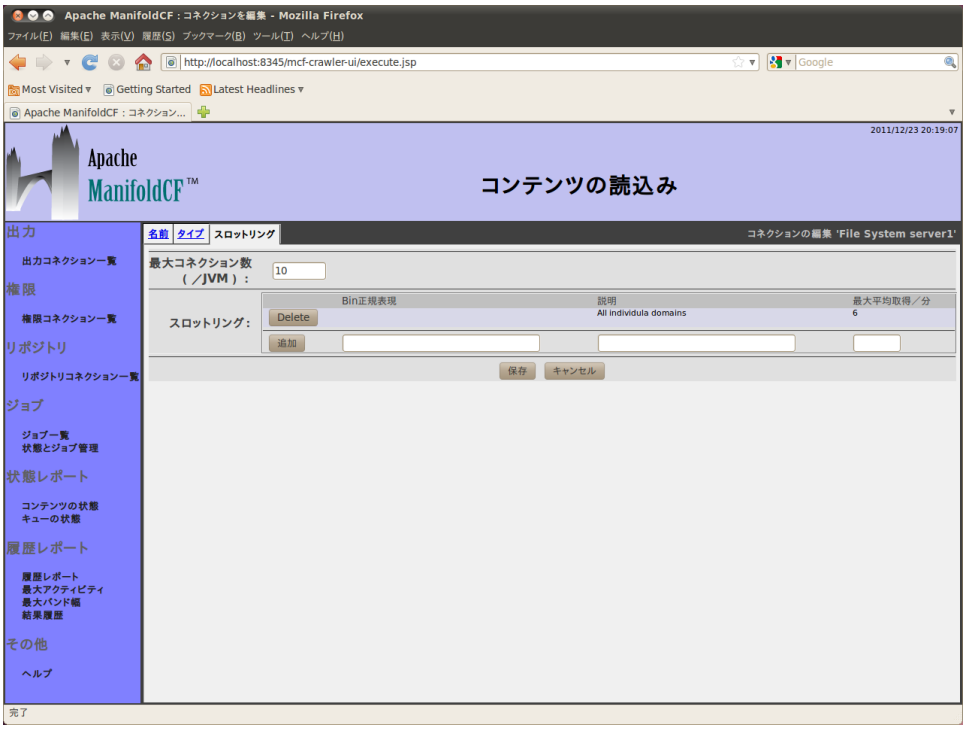
このページには二つの項目があります：先ず一つ目は、システムがこの出力コネクション用に利用できる最大のコネクション数です。システムの負荷を調整したり、ライセンス制約によるコネクション数の制限を行うことができます。値を大きくすると、スループットが上がります。コネクションタイプすべてのデフォルト値は10ですが、出力コネクション・タイプによってはこの値は最適ではない場合もあります。詳細については、以降の各種の出力コネクション・タイプの説明を参照してください。二つ目は、クローラーがこのコネクションを利用して平均的にどれだけ早くコンテンツを読み込むかです。

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。例えば、WebコネクションはHTTPサーバ名毎にスロットルbinを指定することができます。これにより、HTTPサーバ名毎にコンテンツの読み込み頻度を指定することができます。

リポジトリコネクションの「スロットリング」タブから無限のスロットリング設定を定義することができます。スロットリング設定毎にスロットルbin集を表す正規表現、説明文、正規表現毎に一致するスロットルbinの平均読み込み頻度を指定することができます。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、一番保資源を利用しない読み込み設定が有効になります

一番簡単な正規表現は空の式です。この場合はすべてのスロットルbin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事が

できます。平均読込み率を指定して、「追加」ボタンを選択してください。以下のようなスロットリングタブが表示します：



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。

コネクションタイプのタブの詳細に付いては、リポジトリコネクション・タイプによって表示されるタブの説明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ページが表示されます。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「正常」と表示されます。設定に間違いがある場合は、エラー内容が表示されます。エラーメッセージが表示された場合は、設定内容を修正してください。

1.4 ジョブの作成

ManifoldCFの「ジョブ」とは、読込むコンテンツの集まりです。ManifoldCFは指定されたコンテンツの集まりをリポジトリコネクションを介して読込み、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、関連した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、出力コネクションにも依存します。

定義されたジョブの多くは一回以上、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、コンテンツが読込み対象から除外された場合。ManifoldCFはこの両方の場合にも対応しています

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達します。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。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、コンテンツに関連したジョブが無くなります。（注：ManifoldCFジョブは単にコンテンツの読込みタスクではありません。）

ManifoldCFは複数のジョブで1つのコンテンツを読込むことができます。一つ以上のジョブに関連したコンテンツは以下のように処理されます：

- 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情報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最後に実行されたジョブによります。

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、出来る限りこのような状況は避けたほうがよいです。

非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（「シーディんグ」）
2. コンテンツの読込み、新コンテンツの発見、削除対象情報を取得
3. キューから読込み対象外になったコンテンツを削除

ジョブを「継続」に走らすこともできます。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：

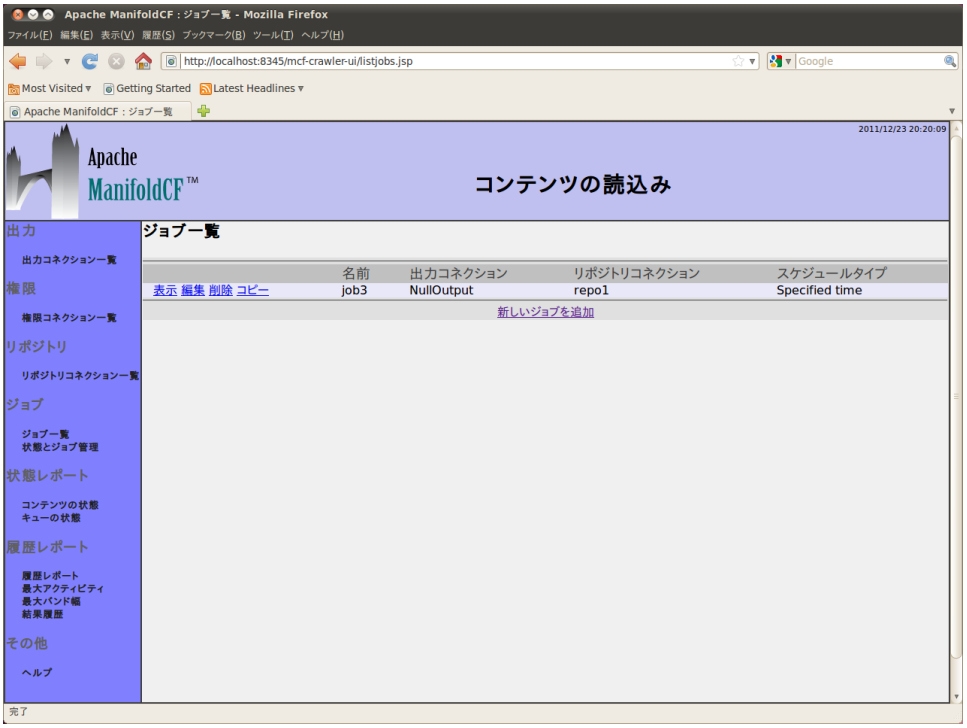
1. ジョブの新規、変更、削除の開始点をキューに登録（「シーディング」）
2. コンテンツの読込み、新コンテンツの発見、削除対象情報を取得。定期的にシードの登録

注：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ません。リポジトリから削除されたコンテンツのみをキューから外す事ができます。

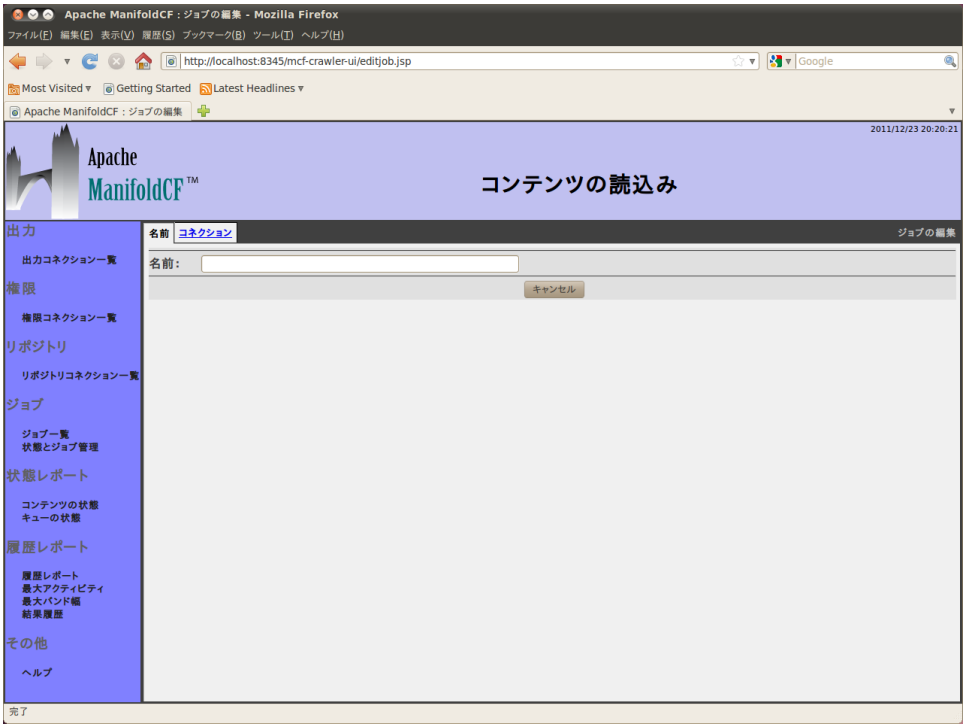
ジョブはユーザが即時に実行することもでき、スケジューリングすることもできます。スケジューリングした場合は、指定日時に開始することも、他ジョブが完了した後に実行するように設定することができます。

平行実行可能なジョブ数の制限は設けられていません。

ジョブを定義する場合は、左メニューの「ジョブ一覧」を選択します。次のようなページが表示されます：



定義されたジョブを表示、編集、削除する場合は、一覧に表示されたジョブの右に表示されるリンクを選択してください。定義したジョブを複写することもできます。新規にジョブを定義する場合は、一覧の下に表示される「新しいジョブの追加」リンクを選択してください。次のページが表示します：



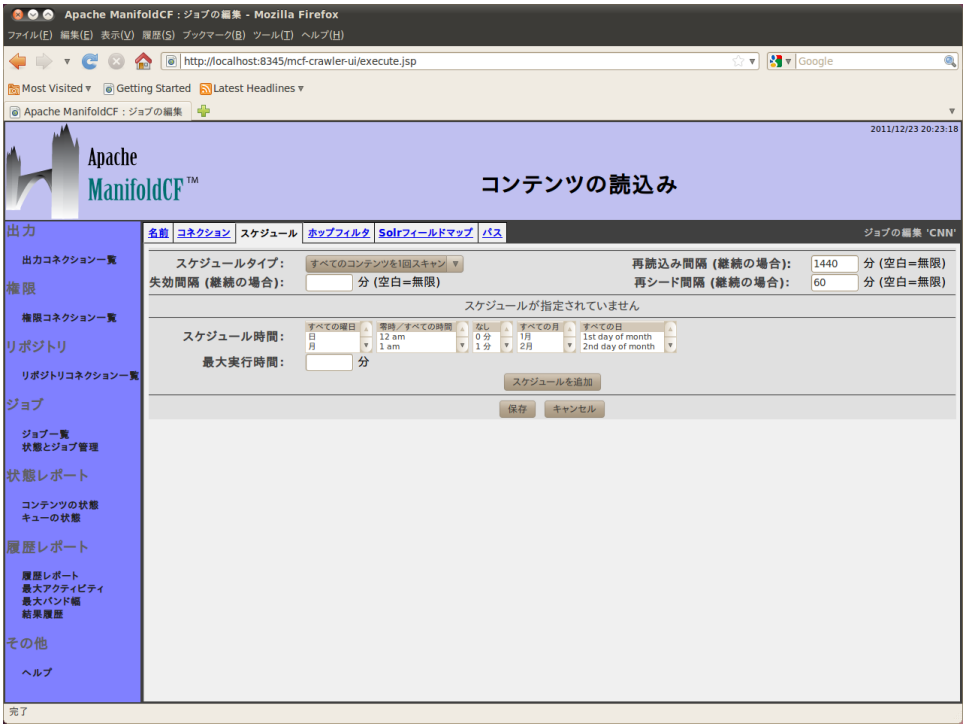
ジョブ名を入力してください。ジョブ名は一意性である必要はありません。ただし、一意性にした方が分かりやすいので一意性にする事を推奨します。入力した後に「コネクション」タグを選択したください：

出力コネクション名と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。

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。ジョブを登録又は更新する場合は必ず「保存」ボタンを押下してください。設定内容を破棄する場合は、「キャンセル」ボタン又は左に表示されるメニュー項目を選択してください。

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うことができます：



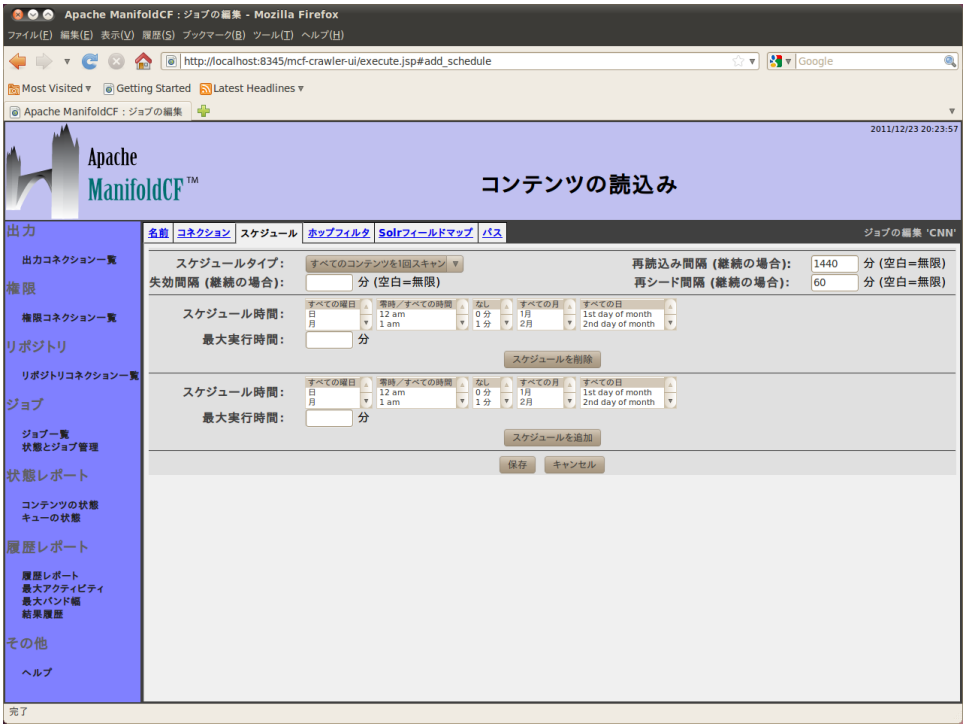
以下の設定を行うことができます:

- ジョブを継続的に実行するか、すべてのコンテンツを一回読み込むか
- コンテンツが無効になるまでの時間。無効になるとコンテンツの索引が削除されます
- コンテンツの更新を確認する間隔
- 初期コンテンツを再シーディングするまでの待ち時間

最後の3つの設定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。各プールダウンメニューから複数の設定を選択された場合は、各プー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:

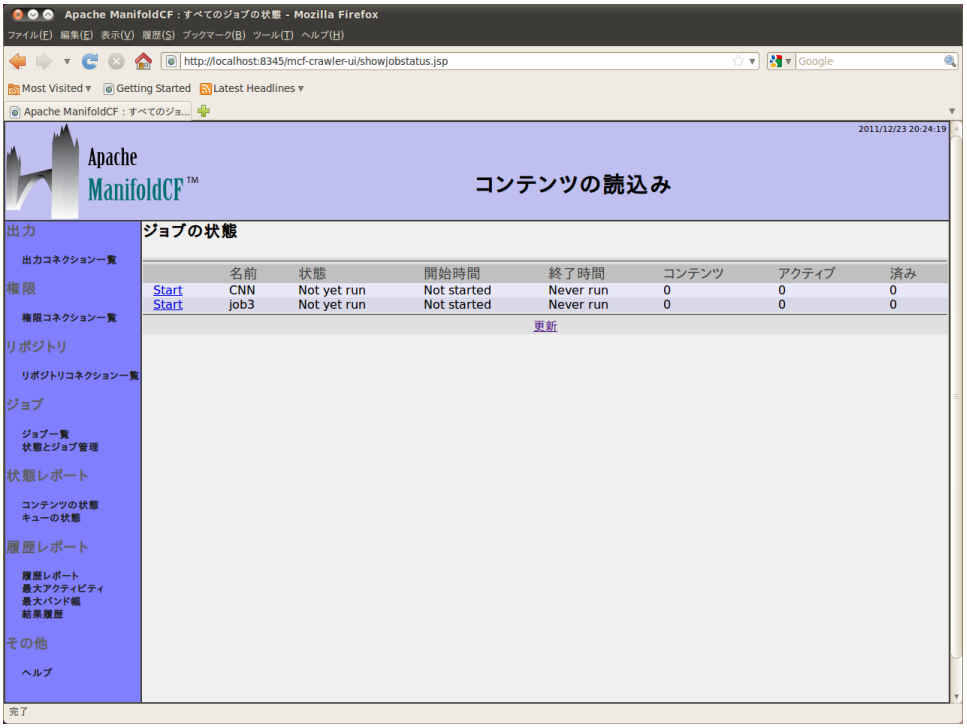


画面例ではジョブを土曜日と日曜日の午前2時から最大4時間（午前6時まで）に実行するように定義しています。

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。

1.5 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「状態と管理」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。ジョブの状態を変更するには、変更するジョブ名の左に表示されている状態のリンクを選択してください。次のような状態処理があります：

- 開始 (ジョブを開始)
- 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実行)

「コンテンツ数」、「処理中」、「処理済み」欄はキューにあるジョブの情報です。「コンテンツ数」はジョブが対象にしているすべてのコンテンツの数です。「処理中」は処理用にキューされているコンテンツの数です。「処理済み」は一回以上はキューされて処理されたコンテンツの数です。

1.6 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。

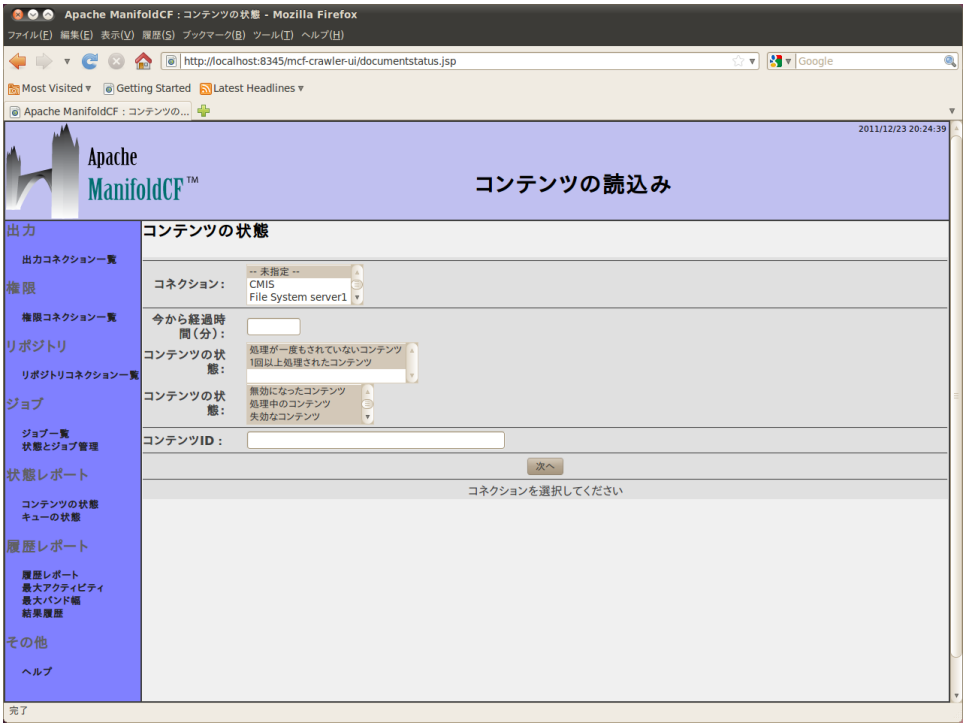
各状態レポートから以下の情報でジョブキューのコンテンツを選択することができます：

- 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

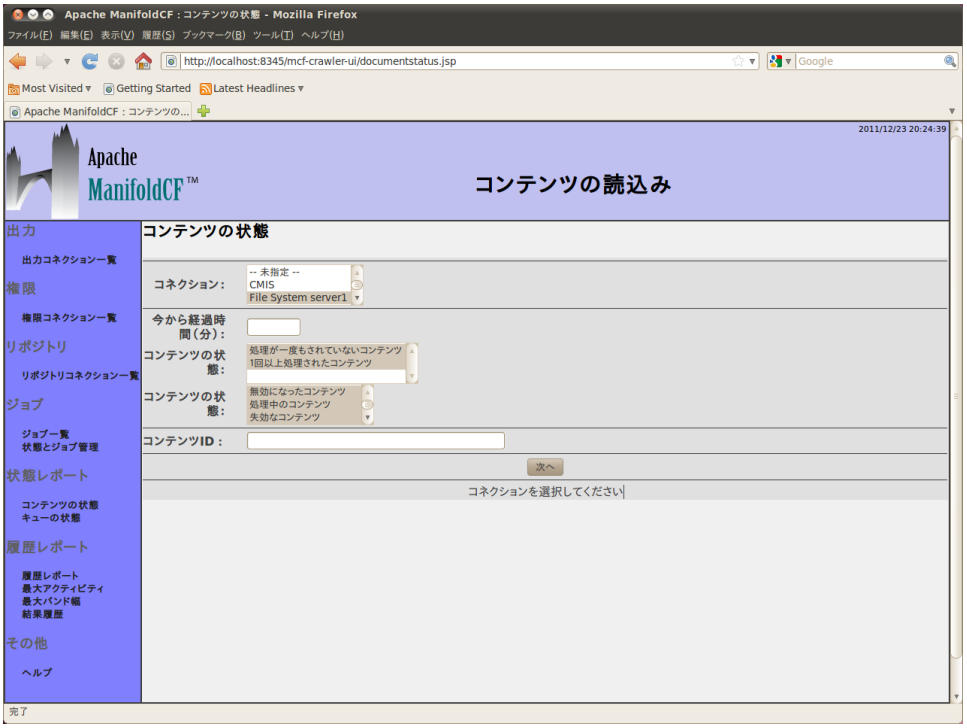
1.6.1 コンテンツ状態

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。

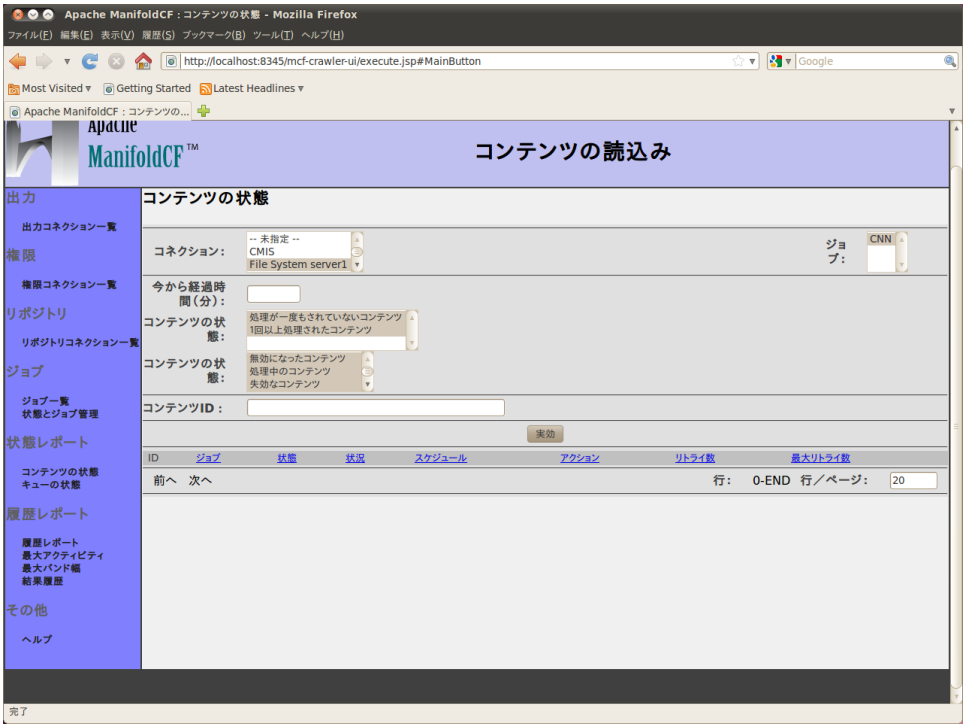
左メニューから「コンテンツ状態」リングを選択すると、以下のようなページが表示されます：



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：



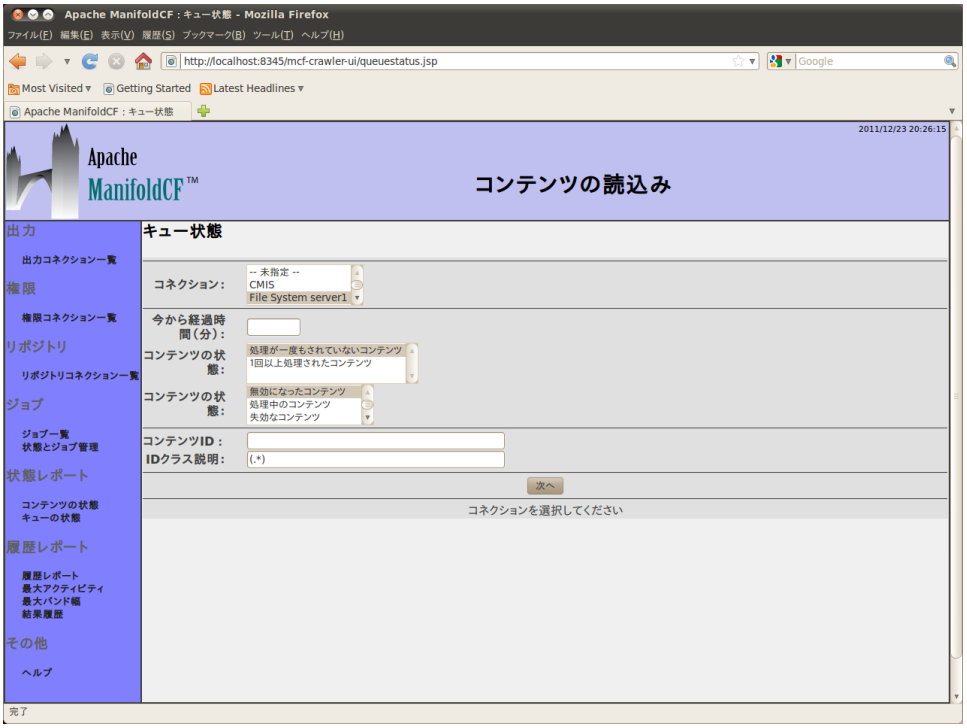
条件を変更して「実行」ボタンを押下して表示するコンテンツ情報を変更することもできます。また、表示する結果数を変更して「実行」ボタンを押下して、1ページに表示するコンテンツ数を変更することもできます。1ページにすべての一致したコンテンツが表示できない場合は、「前へ」リンクと「次へ」リンクを押下した表示する内容を移動することができます。

1.6.2 キューの状態

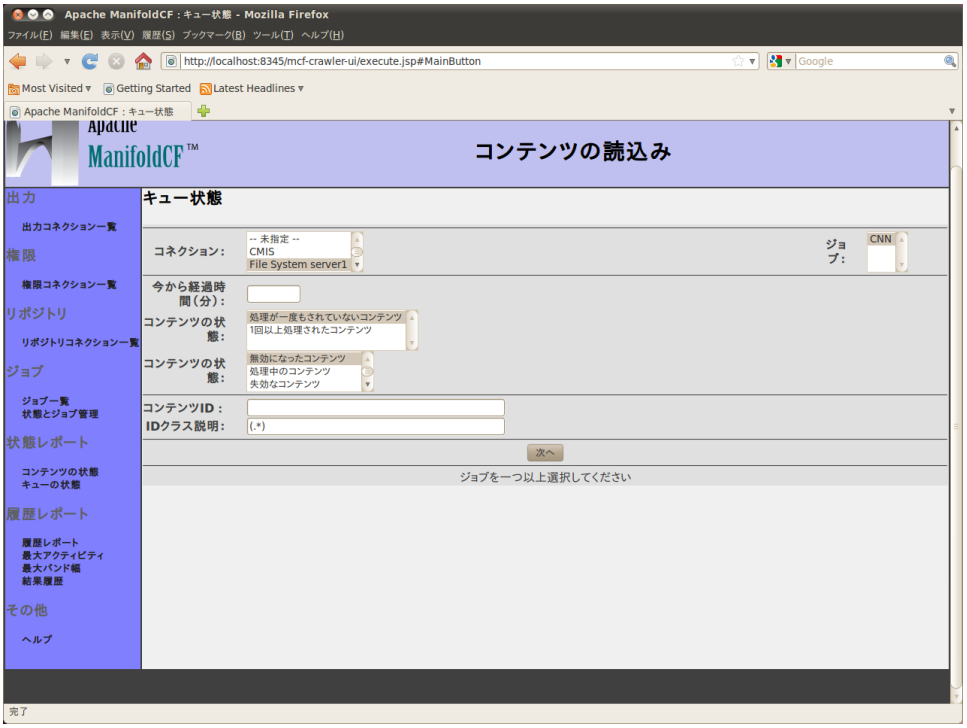
キューの状態レポートは、指定したclassに現れるコンテンツの回数を表示します。classは指定されたコンテンツIDに一致した正規表現のグループとして指定されます。結果はコンテンツの数として表示します。コンテンツの状態と状況の組み合わせ毎に列が設けられます。

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。

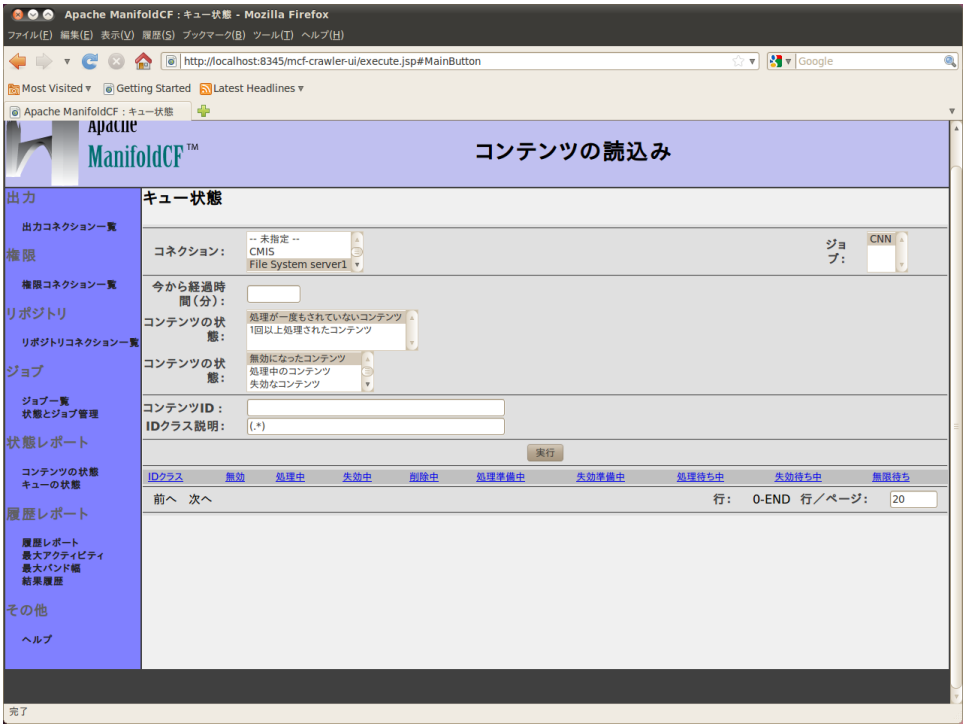
左メニューから「キューの状態」リンクを選択すると、以下のようなページが表示します：



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。

1.7 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
- ジョブを中断
- 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：

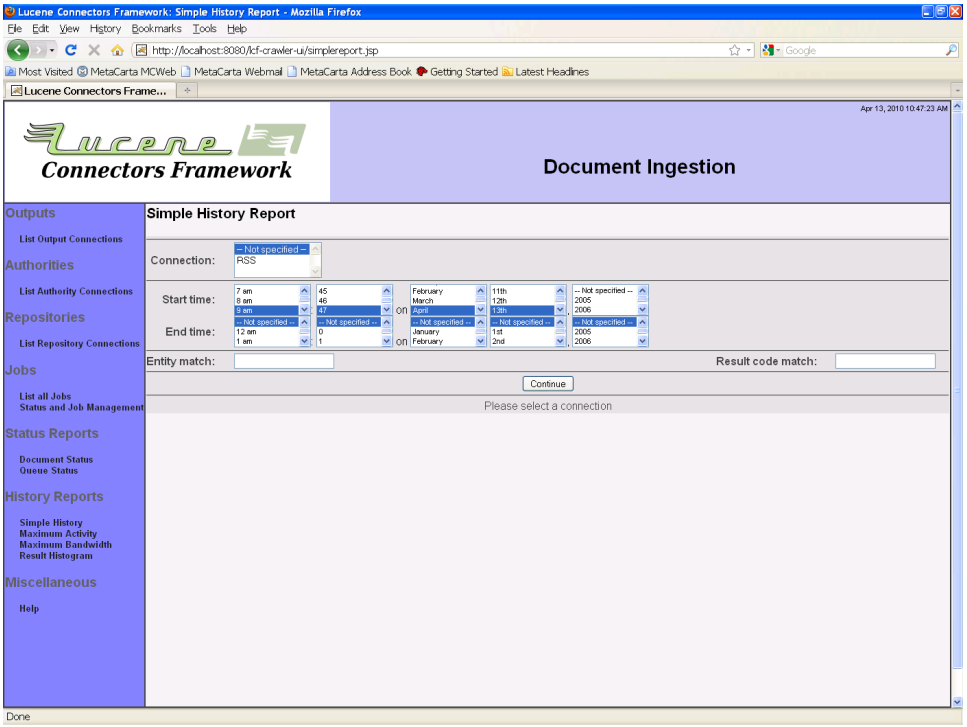
- リポジトリコネクション名
- アクティビティタイプ（複数選択可）
- 開始時間
- 終了時間
- 対象とするコンテンツのID（正規表現で指定）
- 結果（正規表現で指定）

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付いては以下の章を参照にしてください。

1.7.1 簡易履歴レポート

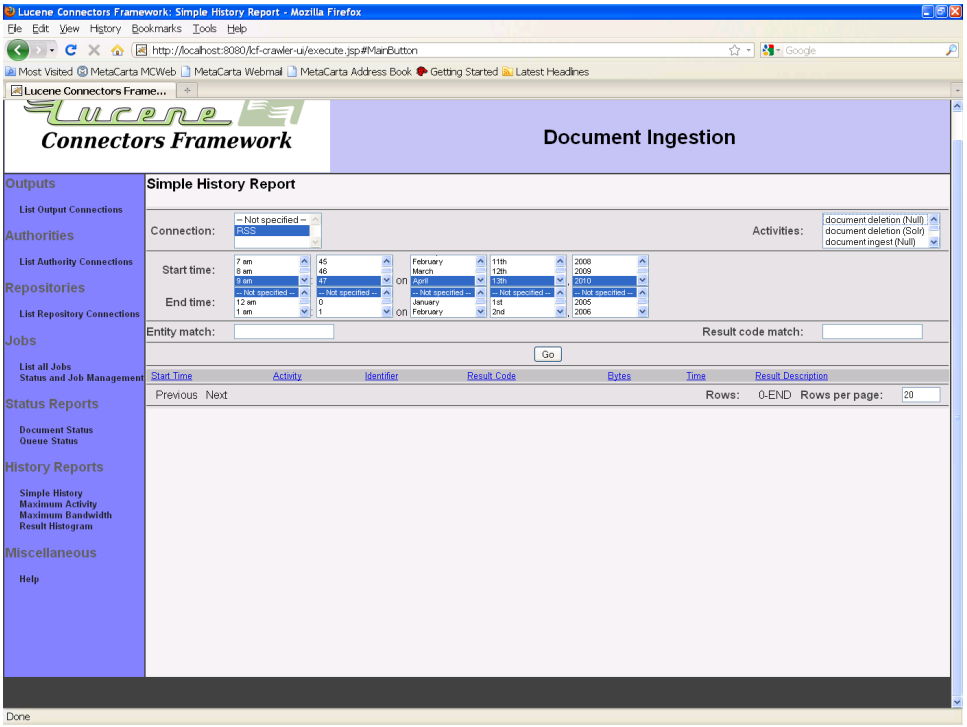
簡易レポートは、集計などは行わずに、条件に一致したリポジトリコネクションの履歴データを表示します。最新イベントから古い順に開始時間、終了時間、処理、ID、データ量（バイト）、結果などが表示されます。表示したレポートのイベント数を変えたり、指定した列順にソートしたり、ページを移動することができます。

左メニューから「簡易履歴」リンクを選択してください。次のようなページが表示します：

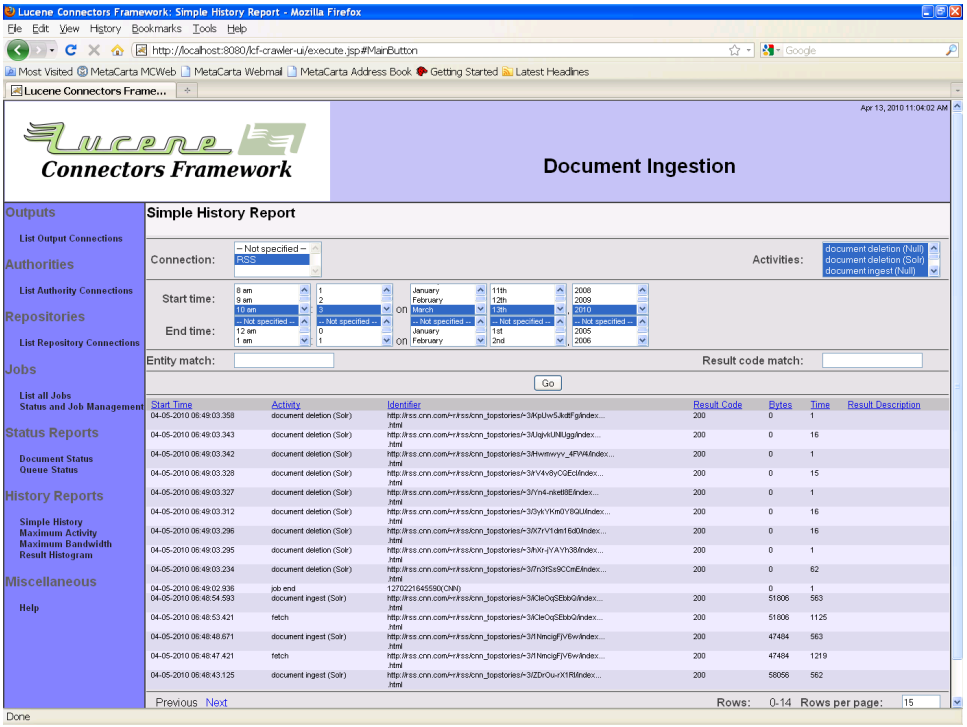


左上のプロダウンメニューからリポジトリコネクションを選択してください。開始と終了日付と時間、対象にするID及び結果コードを正規表現で指定することもできます。デフォルト設定では1時間前のすべてのイベントが選択されます。

「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。



プロダウンメニューから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。

1.7.2 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。

1.7.3 最大バンド幅レポート

最大バンド幅レポートは、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。

1.7.4 結果履歴レポート

結果履歴レポートは、指定したイベントに一致し各結果の数を表示します。

1.8 認証について

選択されたコネクションタイプに認証が必要な場合は、システム管理から必要な情報を入手してください。各コネクションは、コンテンツを読取るのに最低限に必要とする認証で動作するように設計されています。もし実行中にセキュリティに関しての警告が表示した場合は、認証の権限を再確認してください。

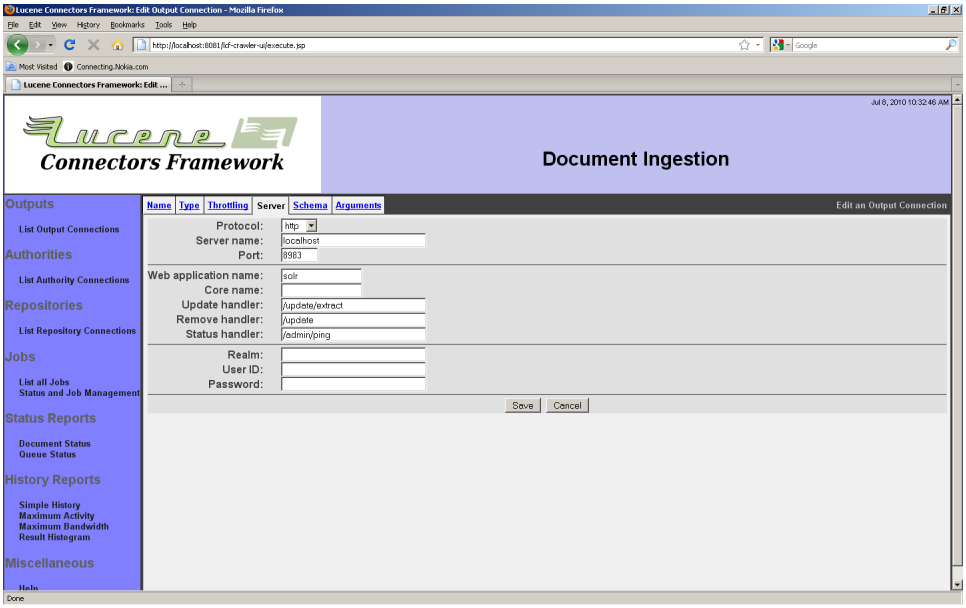
2 出力コネクションタイプ

2.1 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。

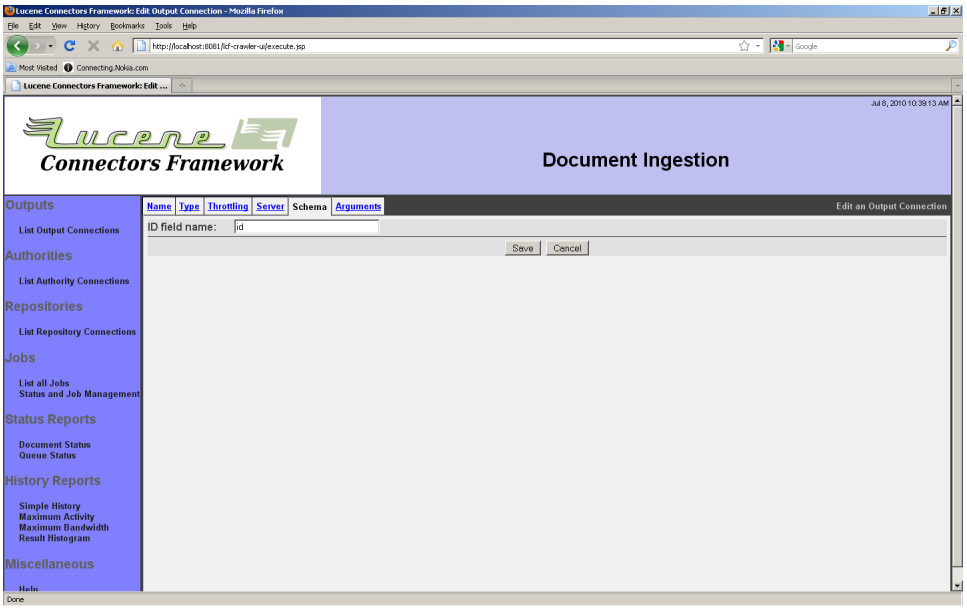
ただし、すべてのコンテンツが送られるため映像のような大きなファイルもフィルタを設定しないと送られてしまい、システムに大きな負荷を掛けてしまいます。不足／間違っている設定を発見してこのような問題を回避するために、Solrコネクションのすべてのクロール結果をレビューすることを推奨します。

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：

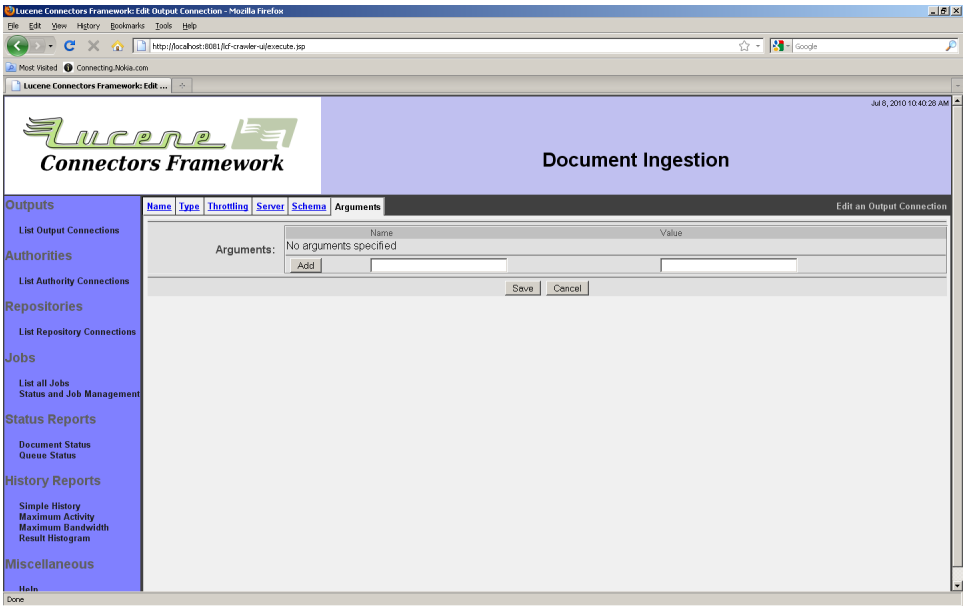


Solrの設定を入力してください。現リリースではベーシック認証のみ対応しています。下の部分にユーザ／パスワードを入力してください。

「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たキーとして使います。

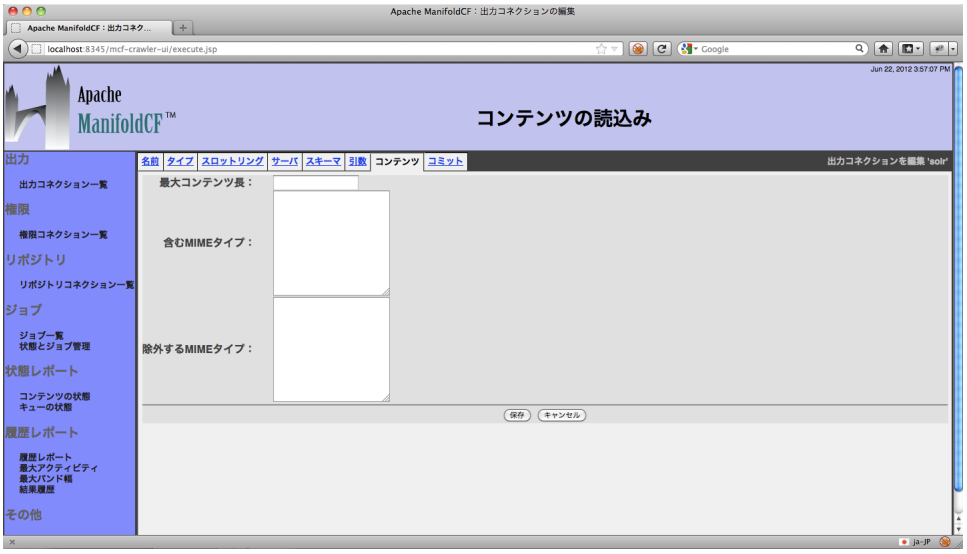


「引数」タブからはSolrに送る任意の引数を指定することができます。Solrの更新リクエストに利用できる任意の引数を利用することができます。 たとえば、Solrのドキュメントを処理するために使われるパイプライン/チェイン：update.chain=myChainを追加することができます。その他に指定可能な引数に付いてはSolrのマニュアルを参照にしてください。タブは以下のように表示されます：

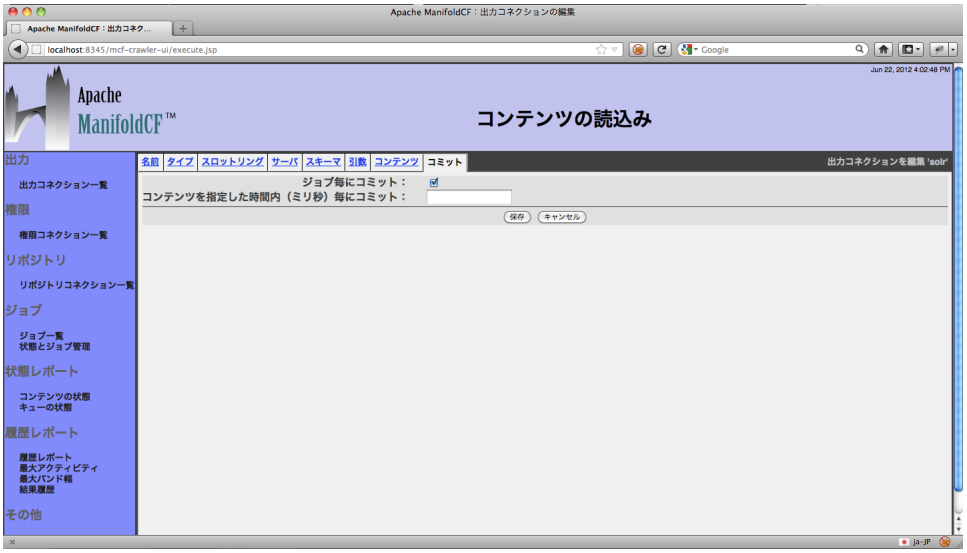


引数名と値を入力して、「追加」ボタンを押下してください。名前が既に存在する場合は、既存の値は新しく指定した値で置き換わります。引数を削除する場合は、削除する引数の左に表示されている「削除」ボタンを押下してください。

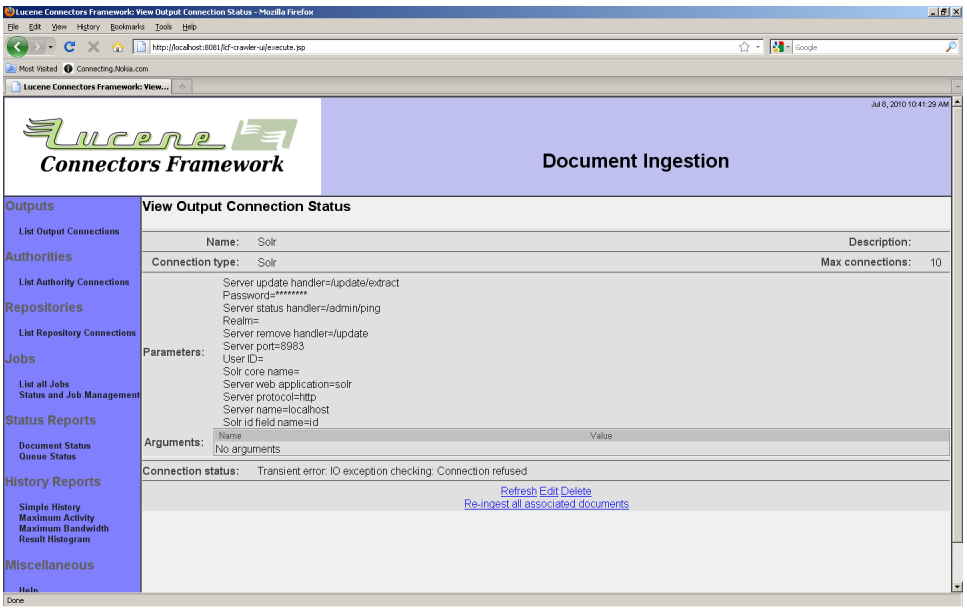
4番目のタブは"コンテンツ"タブです。これはドキュメントのサイズやMIMEタイプに基づいてフィルタリングを行うことができます。ドキュメントのバイト単位の最大長を指定することによって、そのサイズ（例えば10485760は10MBと同じです）を超えたドキュメントを除外することができます。特定のMIMEタイプのドキュメントだけを追加したい場合は、"含むMIMEタイプ"フィールドにそれらを入力することができます（例えばHTML以外のドキュメントを除外するなら"text/html"と登録します）。"除外するMIMEタイプ"フィールドは、特定のMIMEタイプのドキュメントを除外するためのものです（例えばJPEG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



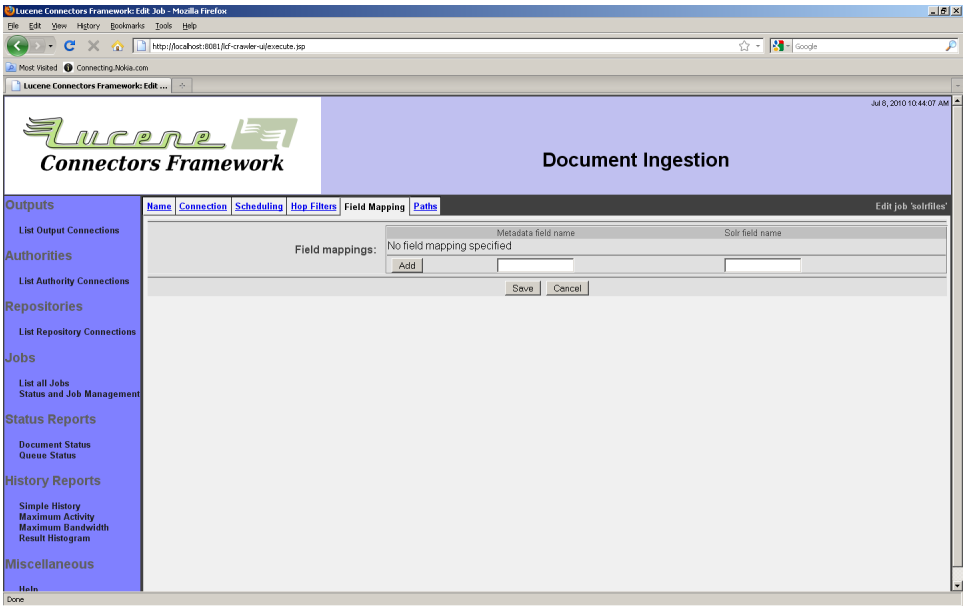
設定の入力を完了した場合は、「保存」ボタンを押下してください。次のような入力した設定一覧が表示します：



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していうます。

ジョブでSolrタイプコネクションを選択すると、「項目マッピング」タブが表示します。このタブからジョブコネクションタイプで取得したメタデータ項目をSolrが受信するように設定された項目と関連付けることができます。メタデータ項目の名前はリポジトリで設定され、Solrス

キーマと不一致の場合があります。指定したメタデータ項目を索引作成の対象外に指定することもこのタブから指定することができます。タブは次のようなページです：

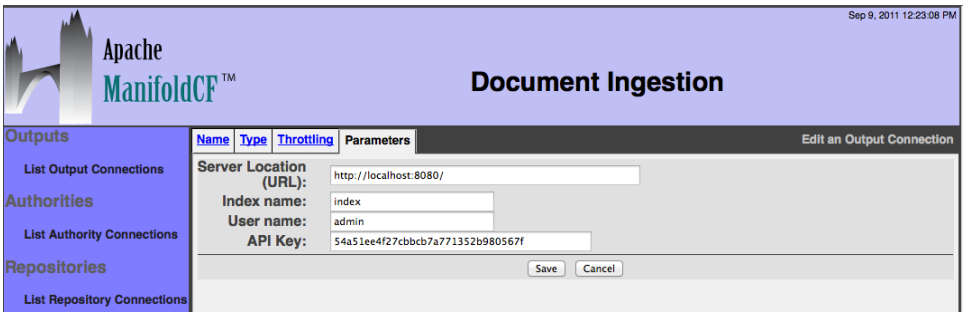


新しいマッピングを追加する場合は、項目「ソース」にメタデータ名、「ターゲット」にSolrの出力項目名を入力して「追加」ボタンを押下してください。Solrに送信しない項目の場合は、「ターゲット」を空に設定してください。

2.2 OpenSearchServer出力コネクション

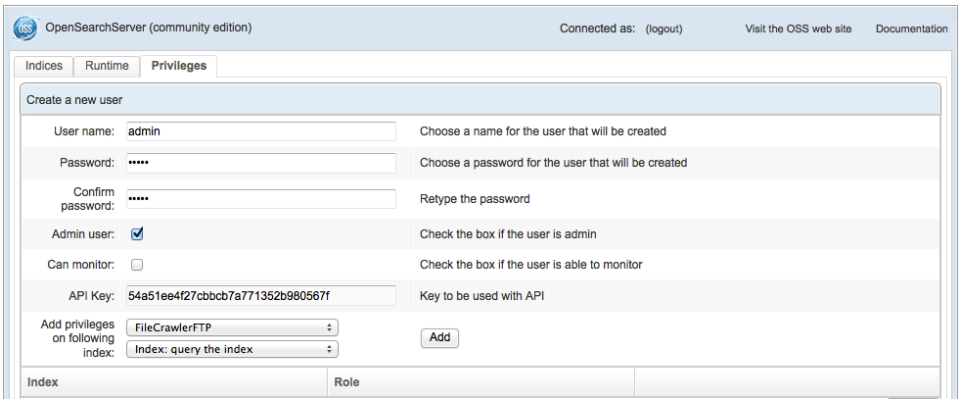
OpenSearchServer出力コネクションは、XML文書をHTTP APIを介してOpenSearchServerに送ります。このコネクタは、できるだけ簡単に利用できるように設計されています。

OpenSearchServer出力コネクションタイプを選択した後には、「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。



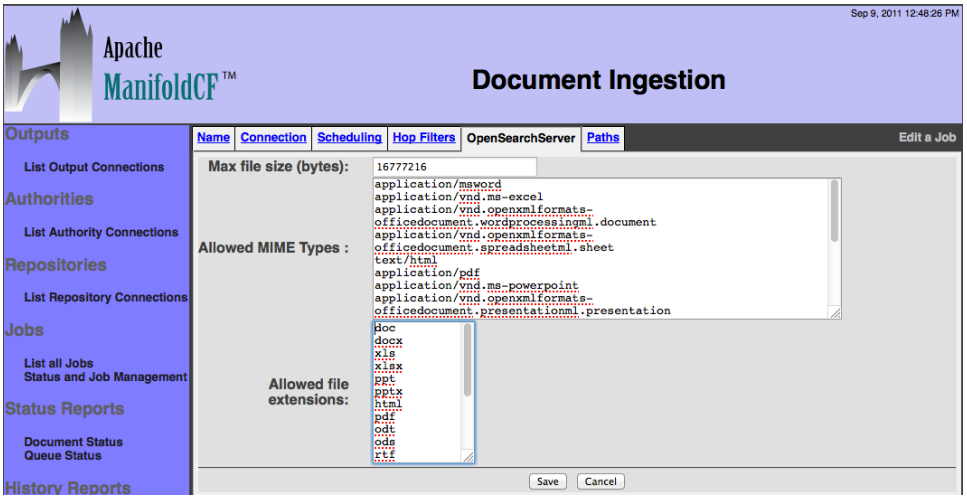
引数:

- 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。

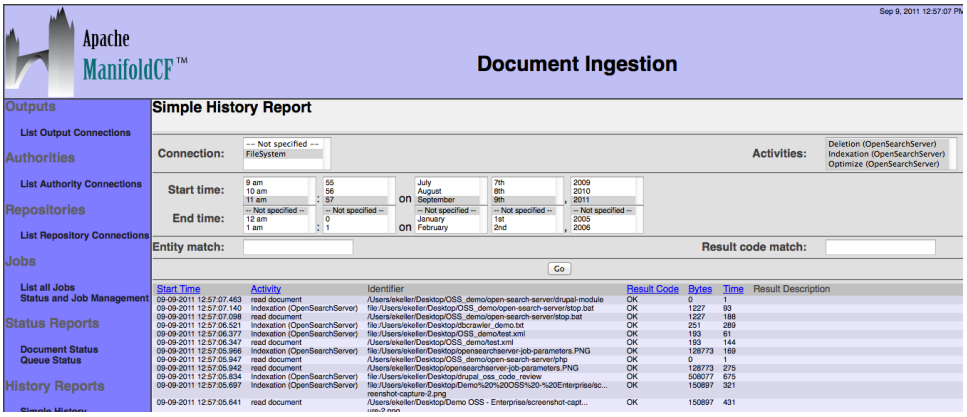


ジョブでOpenSearchServer出力コネクションを選択した場合は、「OpenSearchServer」タブが表示されます。このタブから以下の設定を指定する事ができます:

- 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。



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

2.3 ElasticSearch出力コネクション

ElasticSearch出力コネクションは、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出力コネクションタイプを選択した後には、「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：

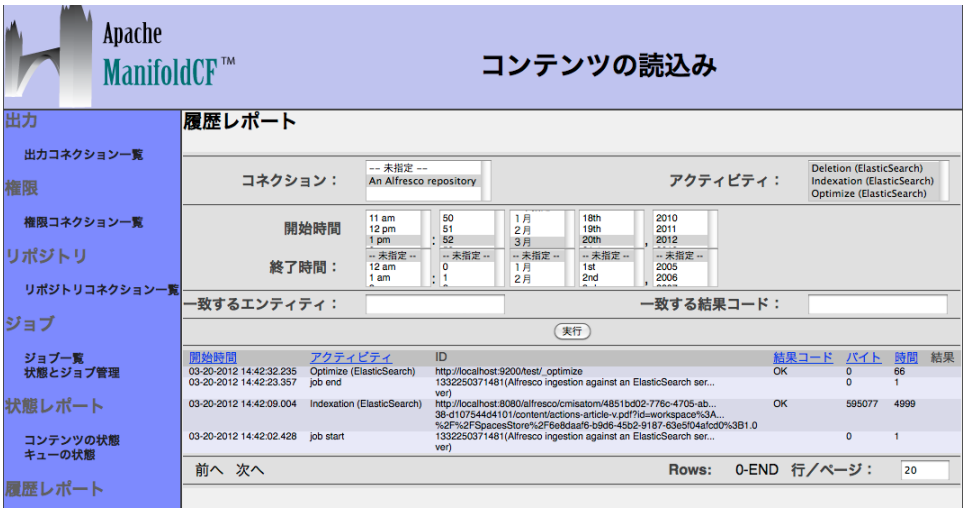
- サーバアドレス：ElasticSearchインスタンスのURL。デフォルトURL（http://localhost:8080）はElasticSearchがManifoldCFと同じサーバで起動している場合のアドレスです。
- 索引名：コネクタは指定された索引にデータを作成します。

- ユーザ名とAPIキー：ElasticSearch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ElasticSearchで、認証情報が記載されている画面です。

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

2.4 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。

2.5 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。

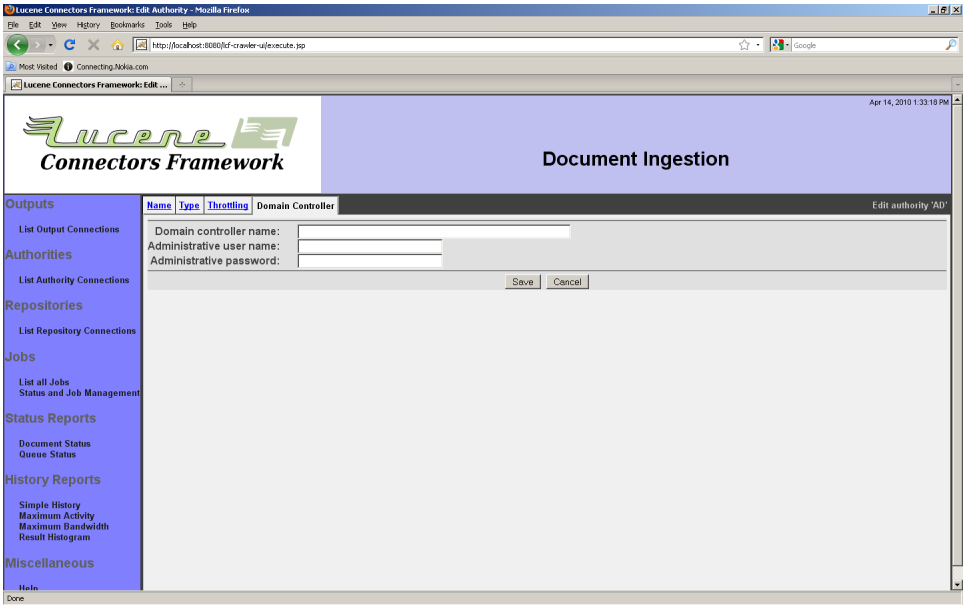
3 権限コネクションタイプ

3.1 アクティブディレクトリ権限コネクション

アクティブディレクトリ権限コネクションは、MS Windows共有ディレクトリ、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。アクティブディレクトリ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:

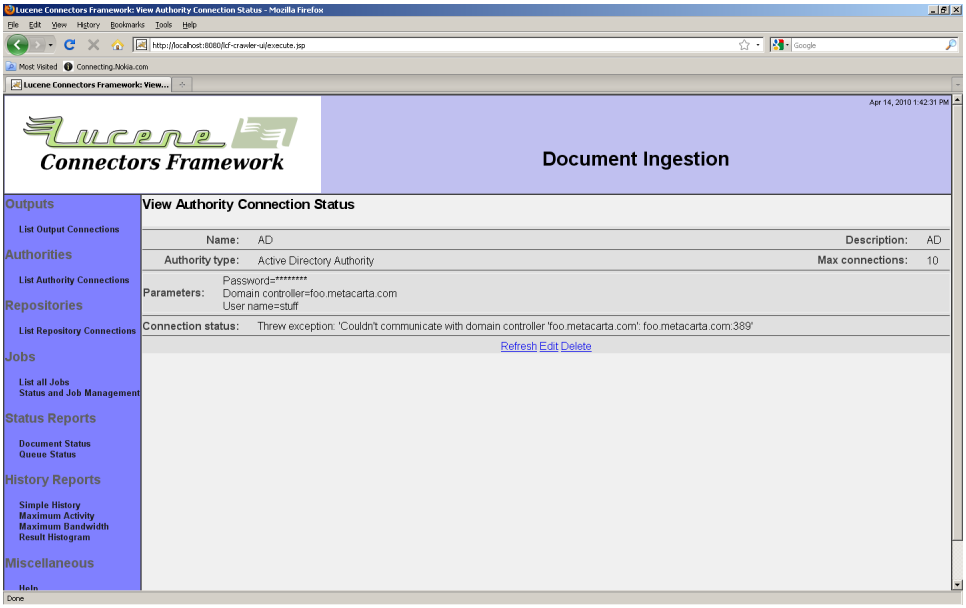
- 子コメントがある場合
- リクエスト／秒が多い場合

アクティブディレクトリ権限コネクションを選択すると「ドメインコントローラ」タブが表示されます:



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します：



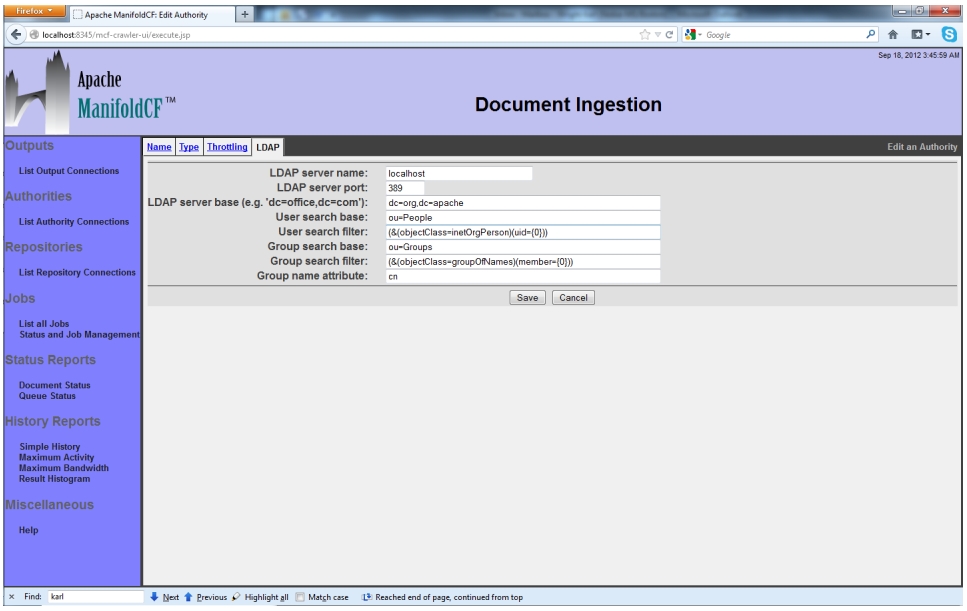
画面ではアクティブディレクトリと接続できないため、エラーメッセージが表示されています。

3.2 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況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。

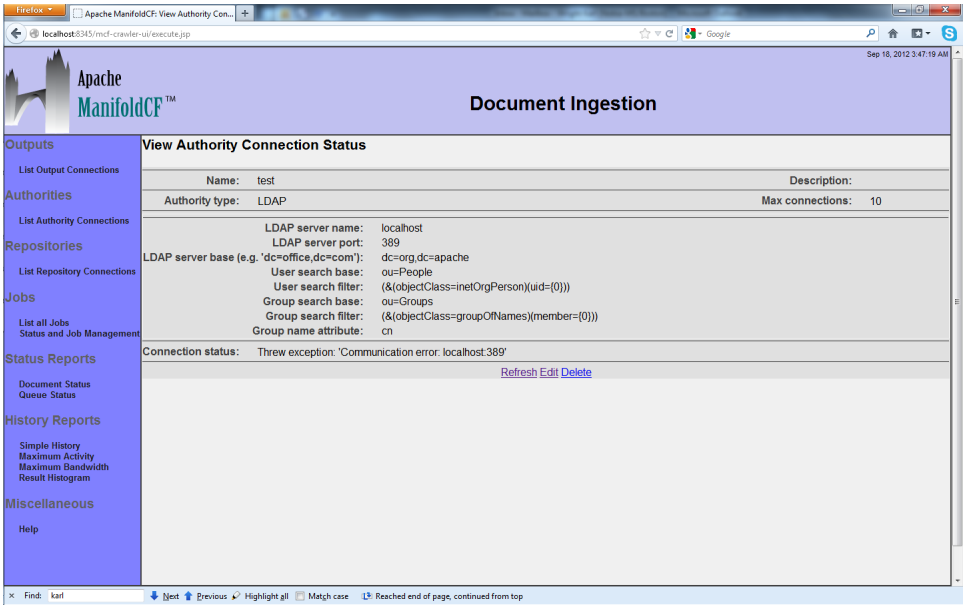
LDAP権限はLDAPサーバーからユーザーまたはグループ名をアクセストークンとして供給することで動作します。これらのアクセストークンは、ジョブごとに入力されたアクセストークンを供給するリポジトリコネクション、またはSamba共有用の、明示的なユーザ/グループ名をサポートしているJCIFSコネクションタイプで使うことができます。

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:



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。

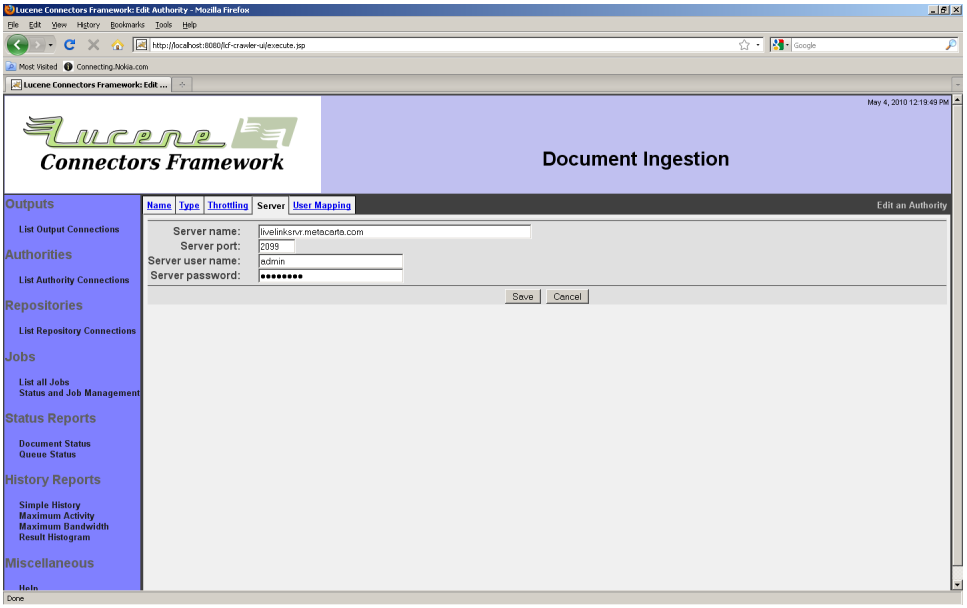
3.3 OpenText LiveLink権限コネクション

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又は正規表現を利用してアクティブディレクトリとLiveLinkユーザと対応付けることもできます。

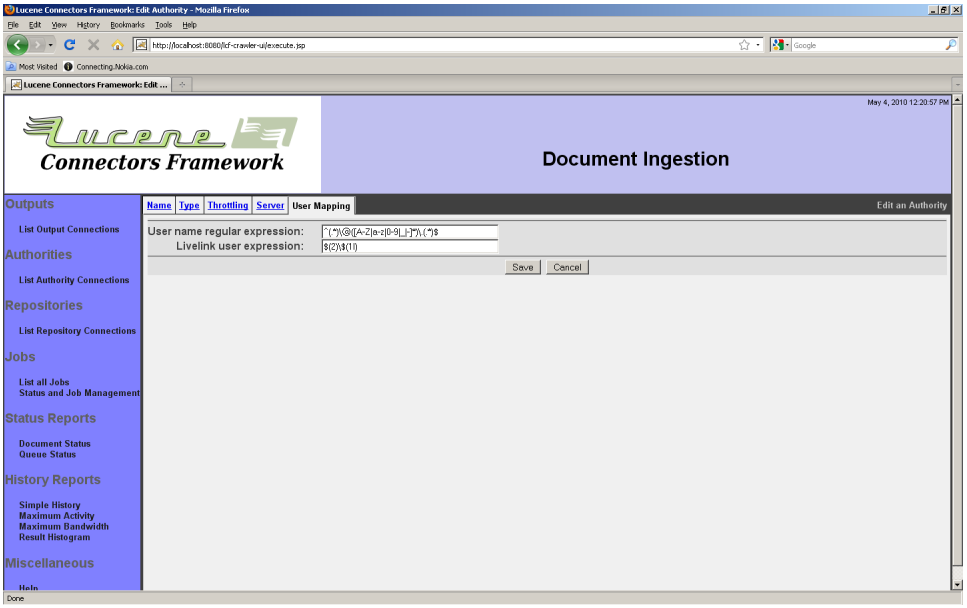
LiveLink権限コネクションを選択すると2つのタブが表示されます:「サーバ」タブと「ユーザマップ」タブ。

「サーバ」タグを選択すると以下のようなページが表示されます:



LiveLinke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します：

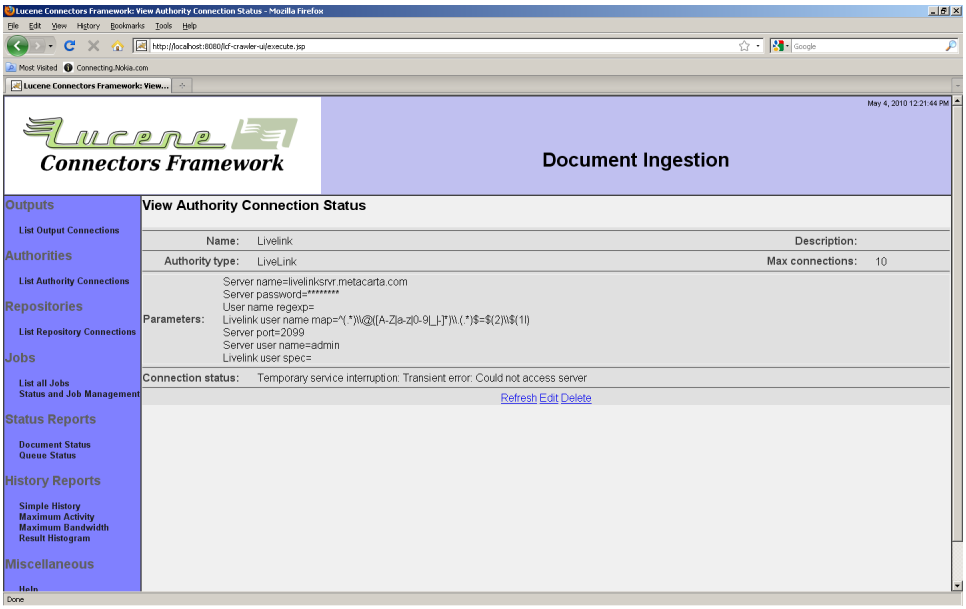


「ユーザマップ」タブから、ユーザ名及びドメイン（通常はアクティブディレクトリから）からの情報をLiveLinkに対応付けることができます。対応は正規表現で定義します。変換元と値は格好（「(」と「)」）で囲みます。括弧に囲まれた部分を「グループ」と言います。置き換え文字列は、固定文字と置き換えグループから構成されます。例えば、「$(1)」は最初に一致したグ

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)はアクティブディレクトリユーザ名を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。

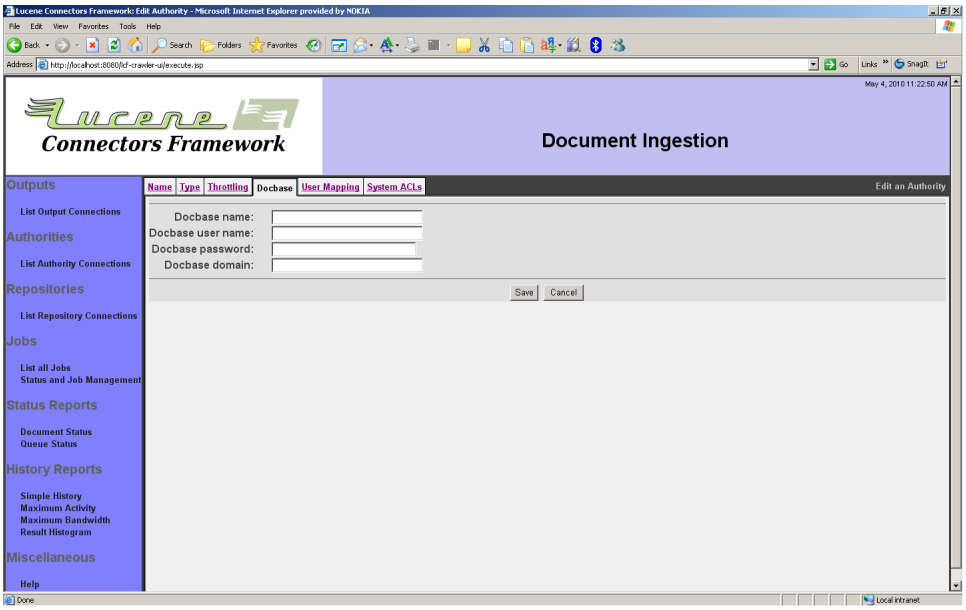
3.4 EMC Documentum権限コネクション

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しまり、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。.

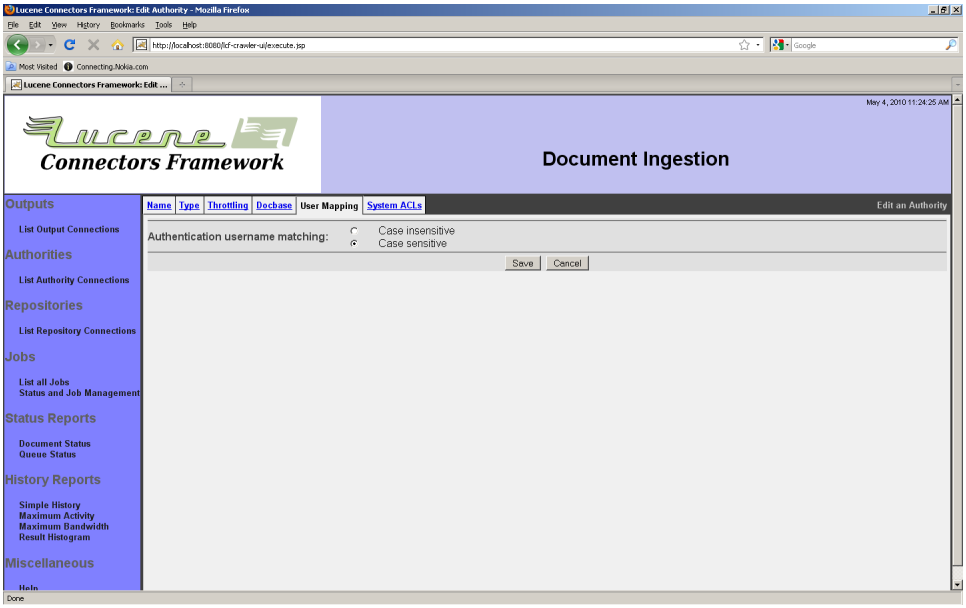
Documentum権限コネクションを選択すると3つのタブが表示します：「Docbase」、「ユーザマップ」、「システムACL」。

Documentum権限コネクションを選択すると、次のような「Docbase」タブが表示します：



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アクティブディレクトリ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。

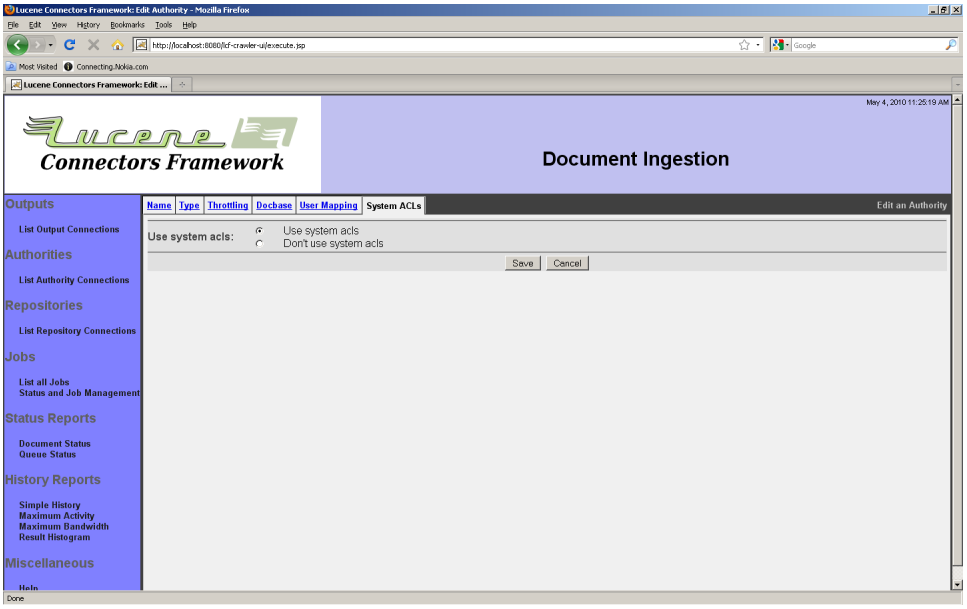
「ユーザマップ」タブを選択すると次のようなページが表示します：



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

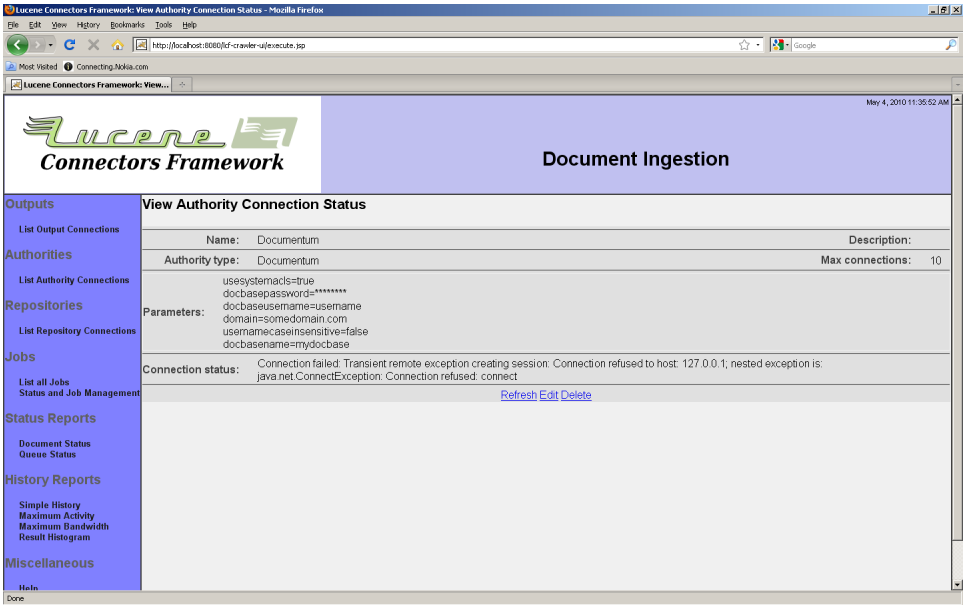
ンスはアクティブディレクトリと連動して、Documentumユーザ名はアクティブディレクトリのユーザ名と同じ、又はアクティブディレクトリユーザ名はただ小文字にされています。詳細に付いては、Documentumシステム管理者ガイドを参照してください。

「システムACL」タブを選択すると次のようなページが表示します：



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。先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：



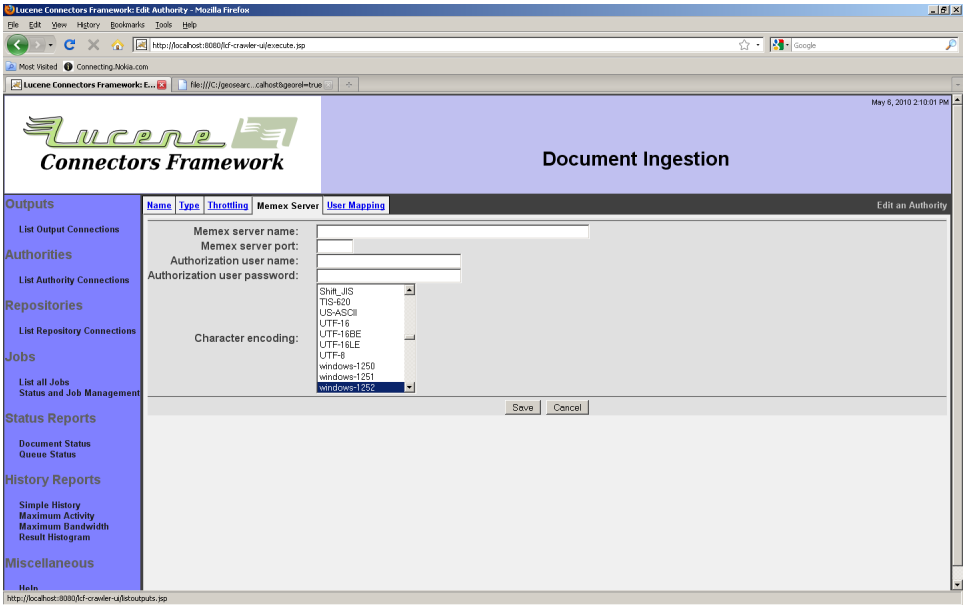
状態を確認して、必要であれば設定内容を修正してください。

3.5 Memex Patriarch権限コネクション

Memem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。

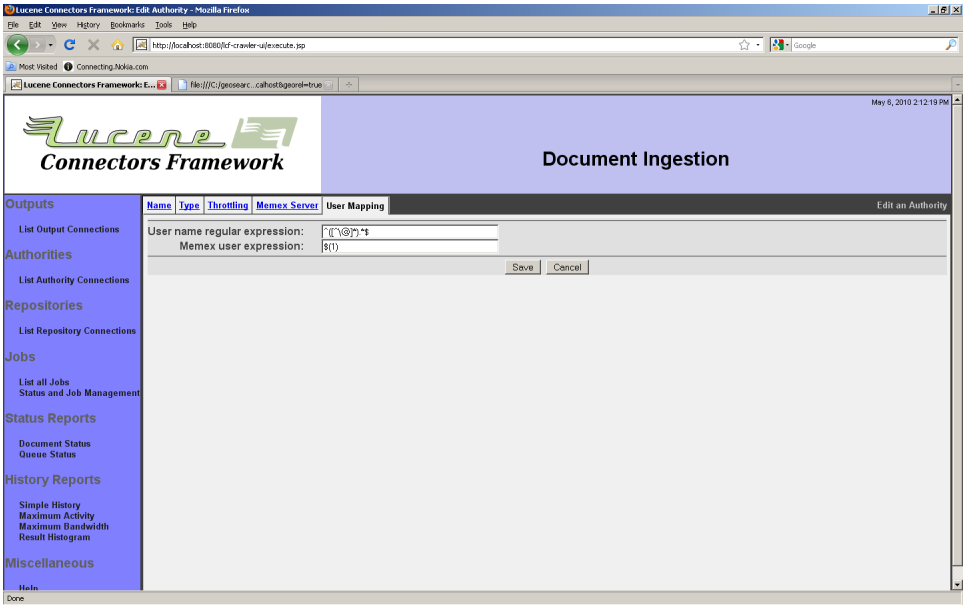
接続するMemexサーバとユーザマッピング情報、Memexサーバからユーザ権限情報を取得するためのユーザの認証情報ををを指定する必要です。

Memex権限コネクションを選択すると、次の2つのタブが表示します：「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します：



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します：

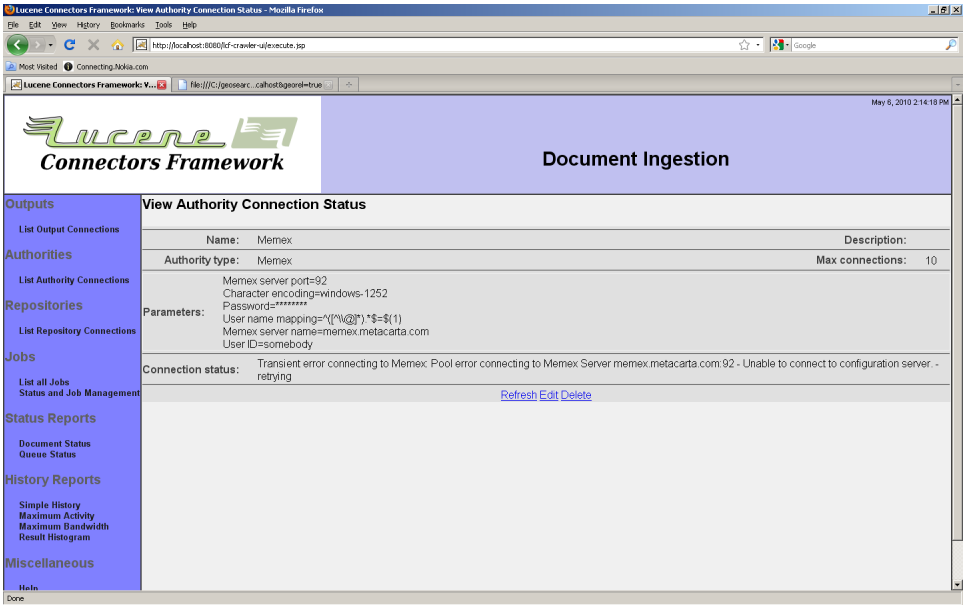


「ユーザマップ」タブから、ユーザ名及びドメイン（通常はアクティブディレクトリから）からの情報をMemexに対応付けることができます。対応は正規表現で定義します 変換元と値は格好（「(」と「)」）で囲みます。括弧に囲まれた部分を「グループ」と言います。置き換え文字

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)はアクティブディレクトリユーザ名を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。

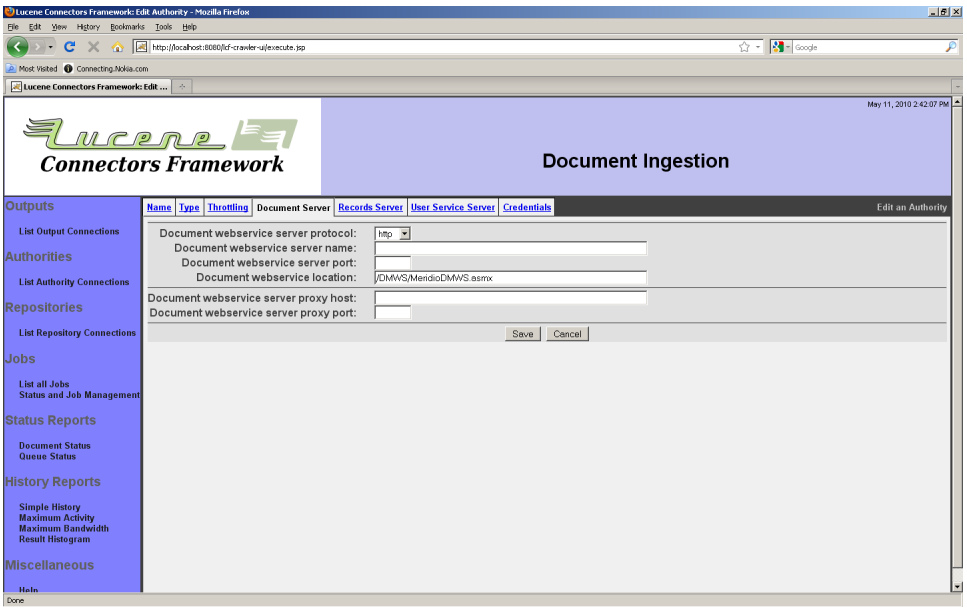
3.6 Autonomy Meridio権限コネクション

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。

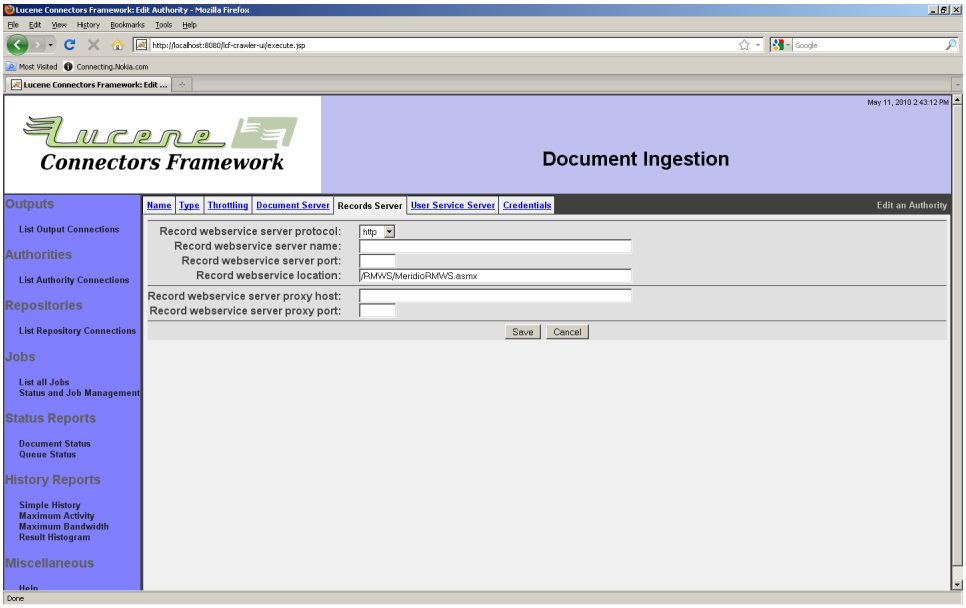
Meridio権限コネクションを選択すると4つのタブが表示されます:「ドキュメントサーバ」、「レコードサーバ」、「ユーザサービスサーバ」、「認証」。「ドキュメントサーバ」タブを選択すると次のようなページが表示されます:



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。プロキシを利用されている場合は、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。

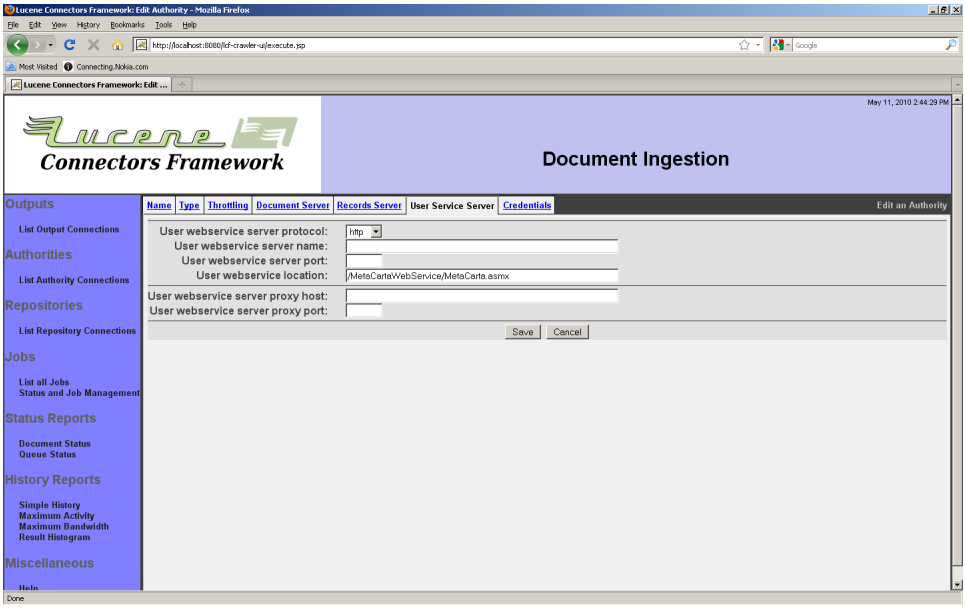
「レコードサーバ」タブを選択すると、次のようなページが表示されます:



プロトコル、サーバ名、ポート番号、Meridioれコートサーバサービスのアドレスを入力してください。プロキシを利用されている場合は、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します：

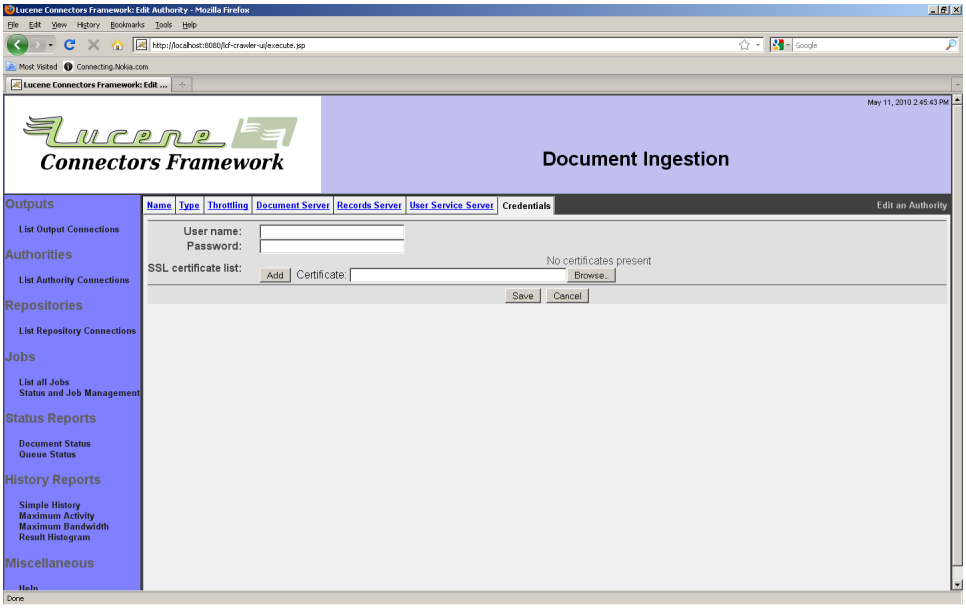


ページ内容を記入するには、Meridio Authorith extensionのインストール先を知る必要があります。

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。プロキシを利用されている場合は、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。

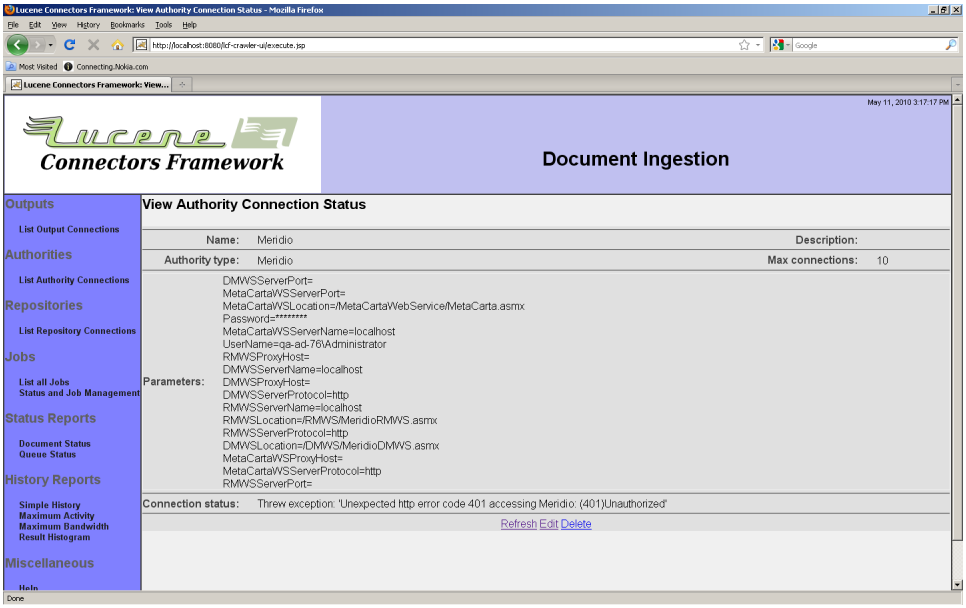
Meridioシステムの場合は異なるサービス毎にサーバを設けることができますが、一般には複数のサービスが同じサーバで動作しています。ただし、コネクションタイプ設定からは異なるサーバを指定することもできます。

「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します：



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）

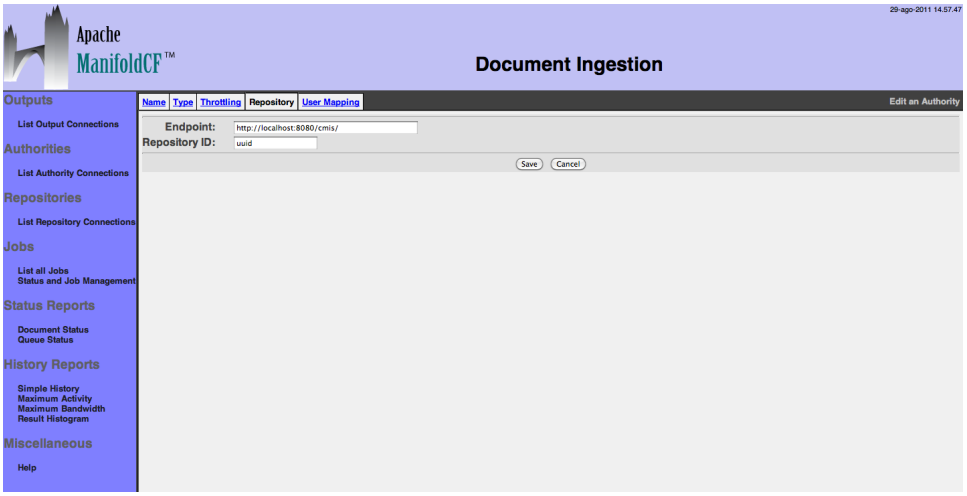
特定のManifoldCFログ情報が必要な場合は、ManifoldCF担当者に連絡してください。

3.7 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、再起表現で指定することができます。

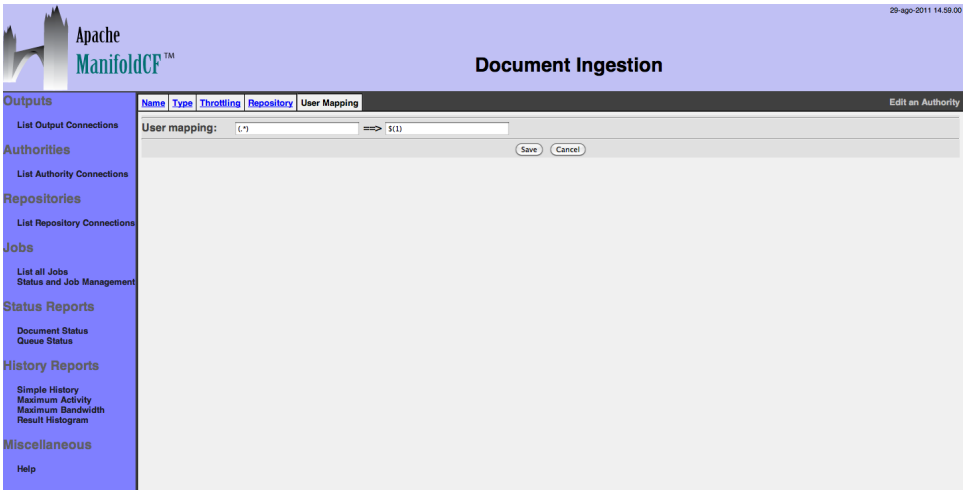
CMIS権限コネクションを選択すると、次の2つのタブが表示します：「リポジトリ」、「ユーザマップ」。「リポジトリ」タブを選択すると次のようなページが表示します：



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。

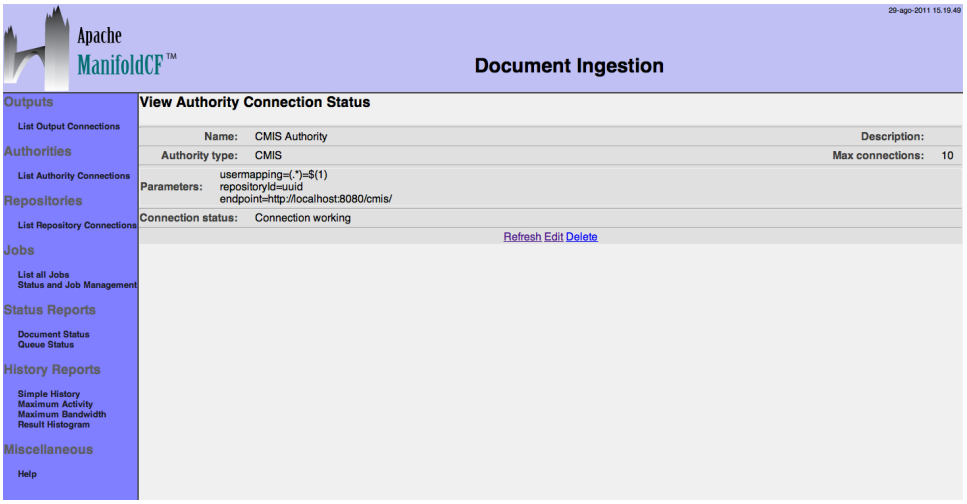
「ユーザマップ」タブを選択すると次のようなページが表示します：



「ユーザマップ」タブから、ユーザ名及びドメイン（通常はアクティブディレクトリから）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。 変換元と値は格好（「(」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)はアクティブディレクトリユーザ名を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：



4 リポジトリコネクションタイプ

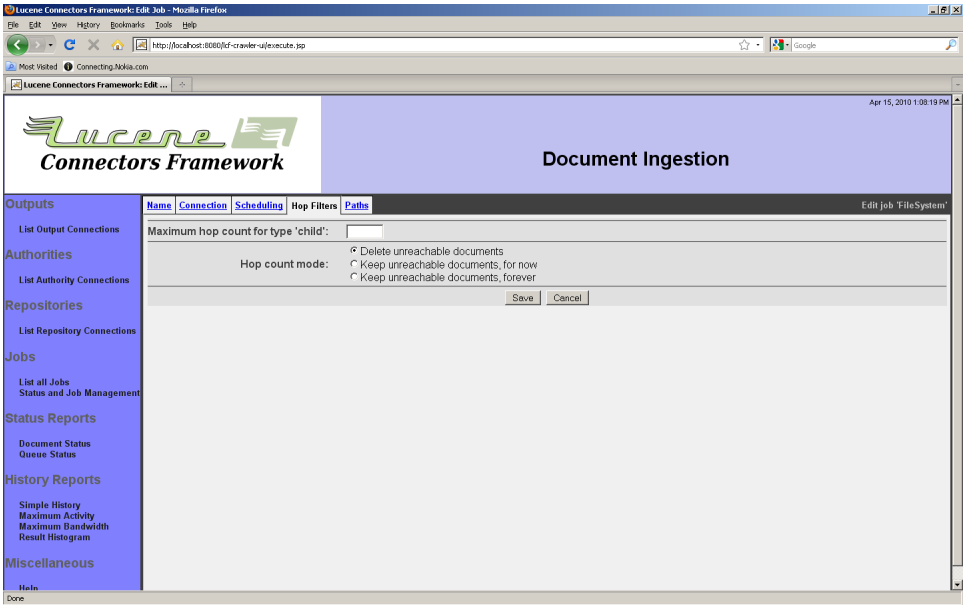
4.1 汎用ファイルシステムリポジトリコネクション

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「スロットルリング」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、又は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

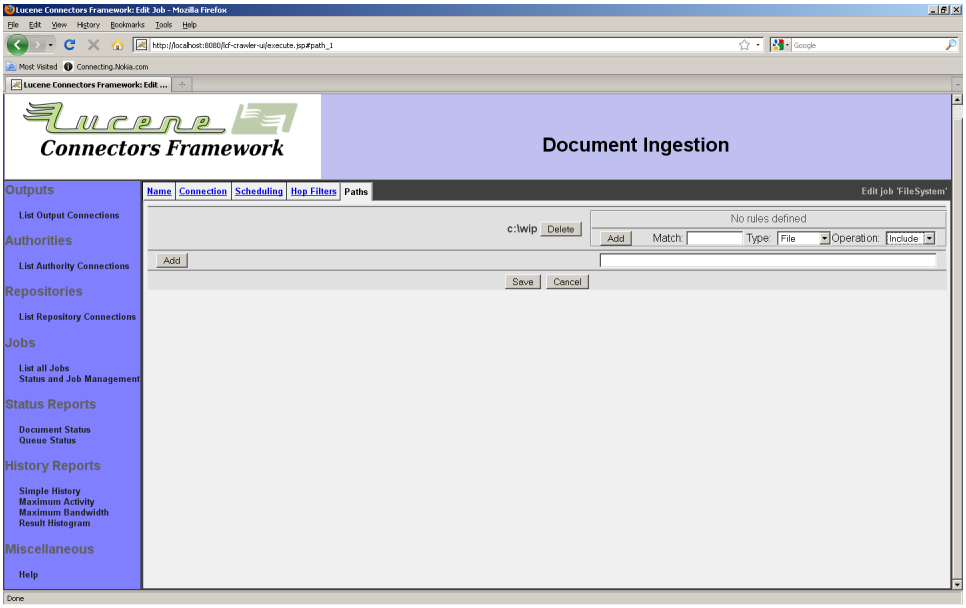
「ホップフィルタ」タブからは、ファイルを取得するサブディレクトリの深さ指定することがでます。ファイルシステムの場合は、使われる場合は少ないかも知れませんが、この設定はWebコネクションタイプでも指定できます。ファイルシステムでこの設定の動作を確認することができます。



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します：



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、ファイル又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。

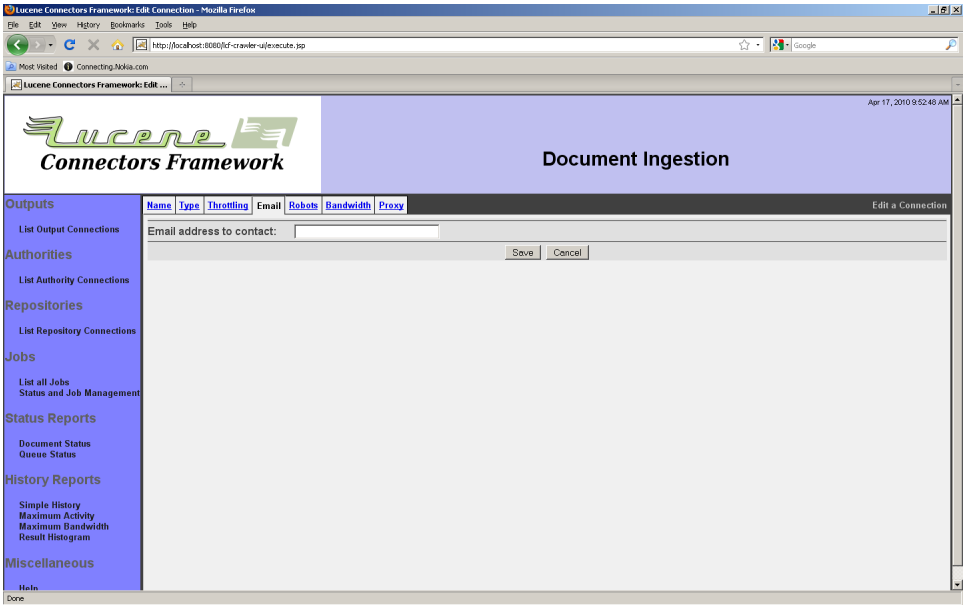
4.2 汎用RSSリポジトリコネクション

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、通常のコンテンツを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、コンテンツを再読込みしないように設定し、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。

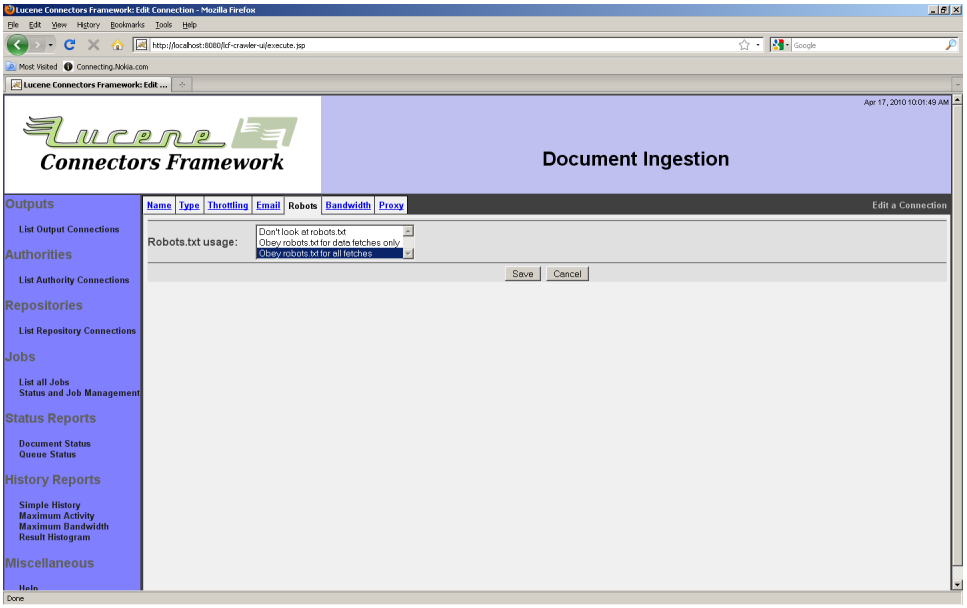
RSSコネクションには4つの固有タグがあります:「メール」、「ロボット」、「バンド幅」、「プロキシ」。「メール」タブを選択すると以下のようなページが表示します:



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。

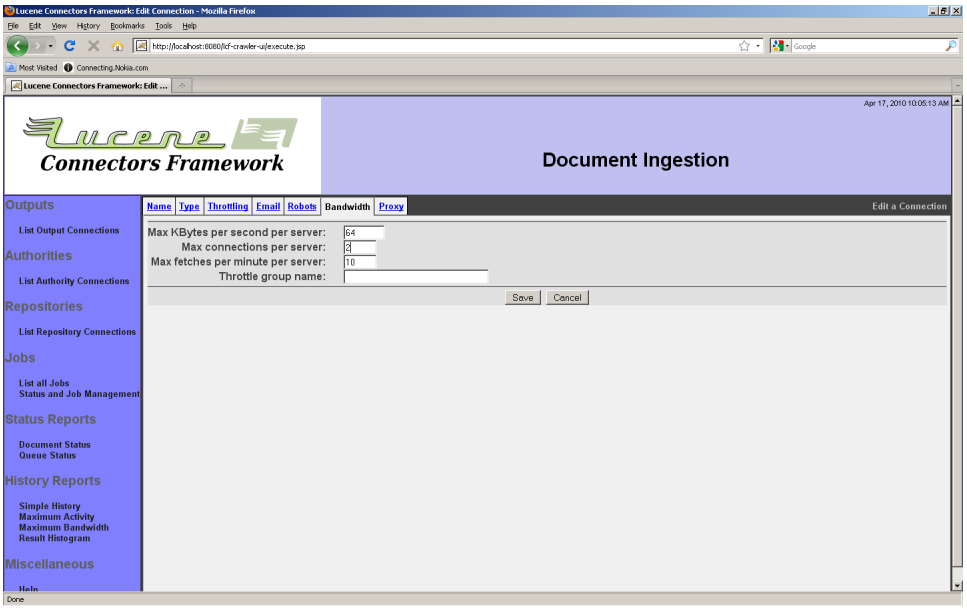
この項目は必須です。RSSコネクションはメールアドレスは妥当性を検証しませんが、ウエブ住民として正しいアドレスを入力してください。なお、サーバ管理者は「悪い」リクエストを拒否するように設定を変えることができますので、相手側サーバのことも考えてスロットリング設定を行ってください。

「ロボット」タブを選択すると次のようなページが表示します:



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します：



サーバ毎に、コネクション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コネクション数も指定できます。

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。

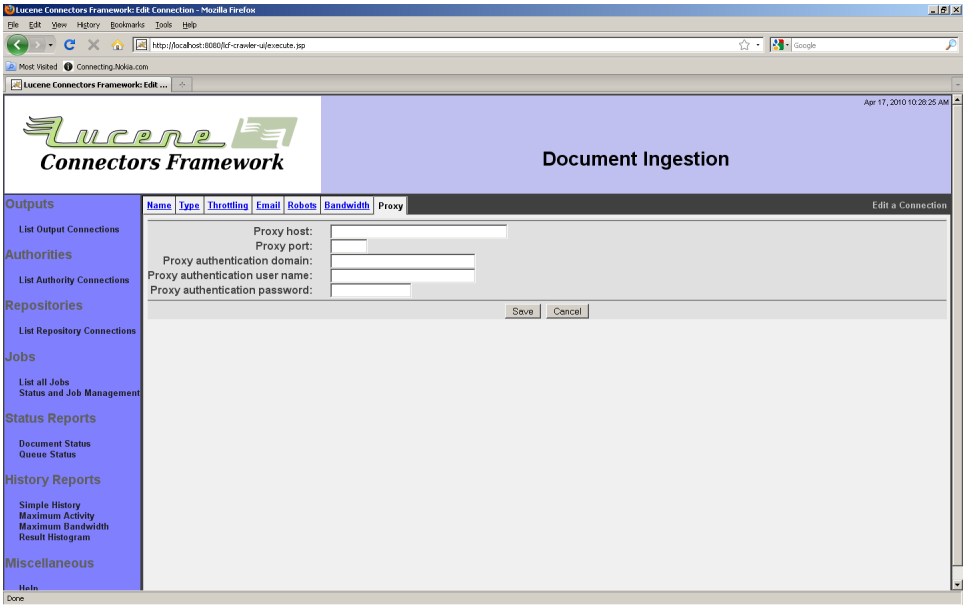
「スロットリンググループ」は、複数のRSSタイプコネクションのスロットリング設定を1つに纏めるための設定です。するっとリンググループ名が同じRSSタイプコネクションは同じスロットリングプールに纏められます。

「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。

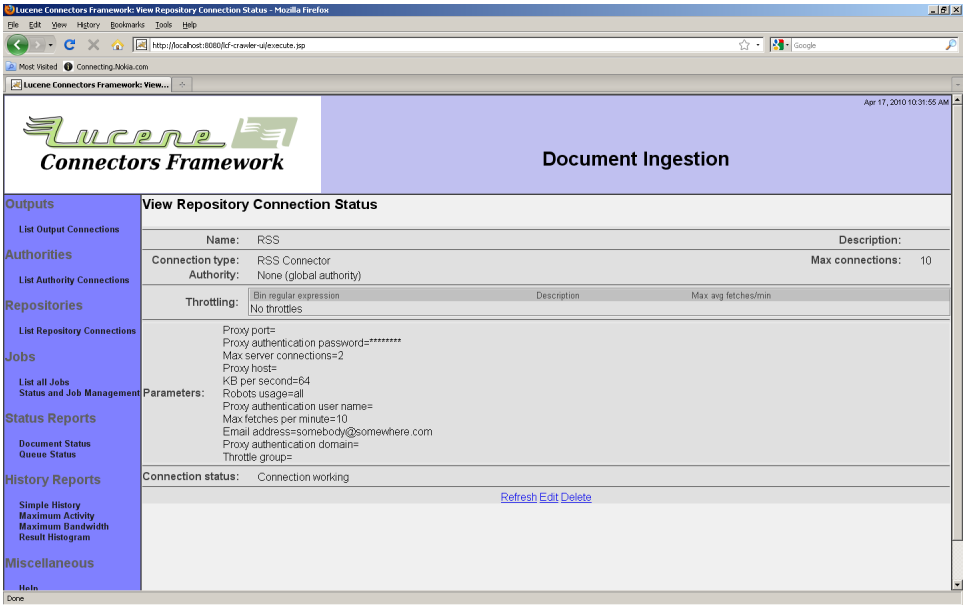
そのような理由のため、RSSコネクションには「バンド幅」タブと「スロットリング」タブの両方を設定することを推奨します。最大転送率を「バンド幅」タブで設定し、平均転送率を「スロットリング」タグで設定します。RSSコネクションのコンテンツIDはコンテンツのURLです。URLのbin名はサーバ名です。なお、「最大コネクション数／JVM」はデフォルトでは10です。この値は、RSSコネクションタイプには最適ではない可能性が高いです。ワーカスレッド毎にコネクションを1つ設けることを推奨します。デフォルトではワーカスレッド数は30ですので、「最大コネクション数／JVM」を30に設定にすることを推奨します。

プロキシを利用されている場合は、「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します：

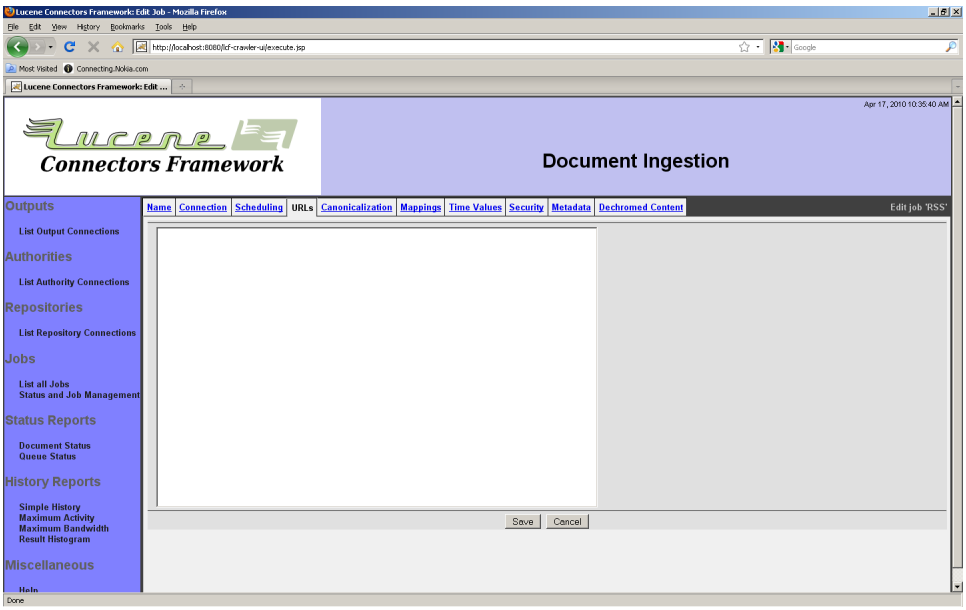


「プロキシホスト」にはプロキシサーバアドレスを入力してください。「プロキシポート」にはプロキシのポート番号を入力してください。認証が必要な場合は、ドメイン名、ユーザ名、パスワードを入力してください。プロキシを利用されない場合は、プロキシ関連のすべての項目を空にして置いてください。

情報を入力した後に「保存」



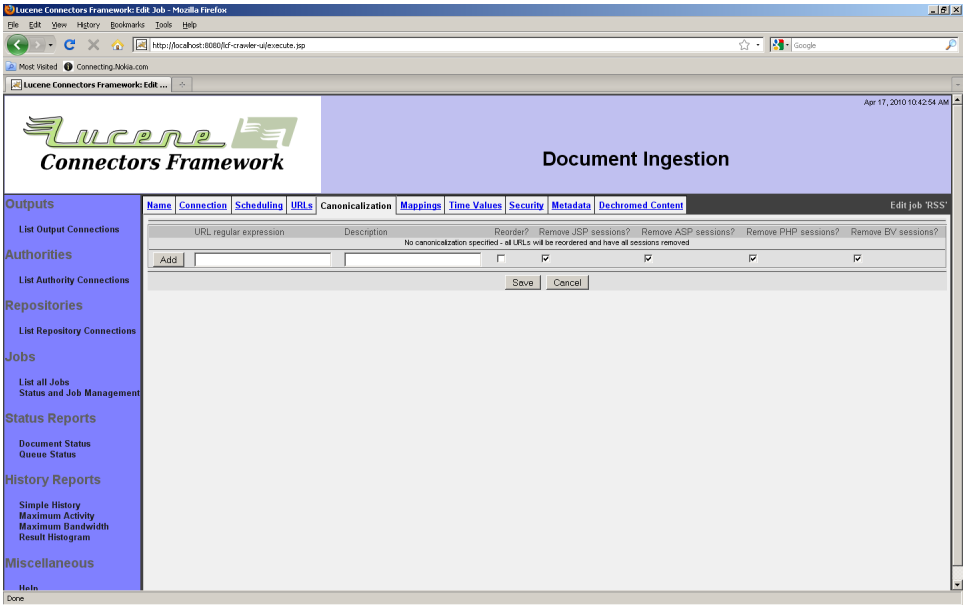
ジョブでRSSコネクションタイプを選択すると、次のタブが表示されます:「URL」、「正規化」、「マッピング」、「時間」、「セキュリティ」、「メタデータ」、「索引対象」。「URL」タブからジョブで対象にするRSSフィードの情報を指定します。「URL」タブを選択すると次のようなページが表示します:



読込む意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い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すです。その他にもURIにセッションクッキ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します:

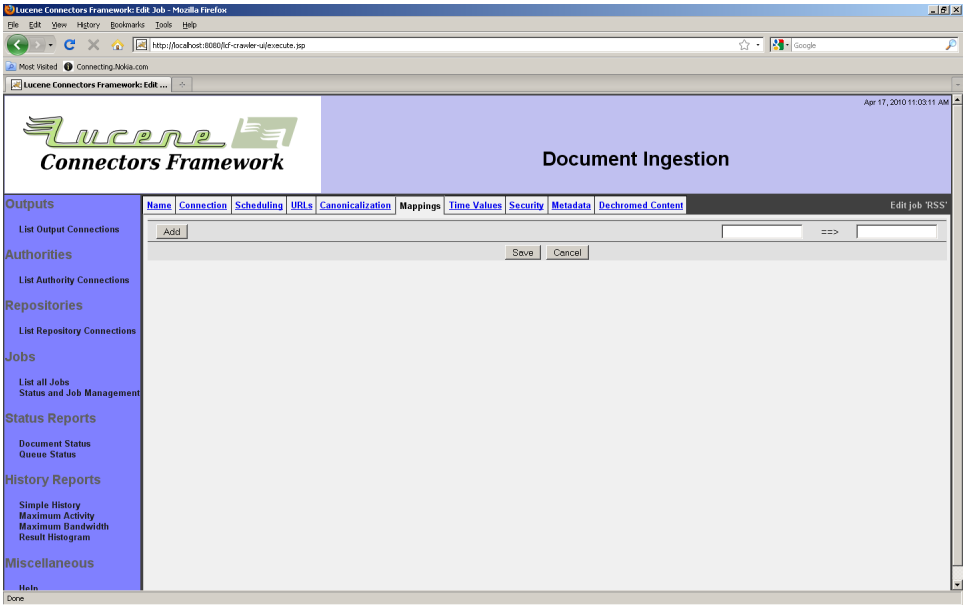


正規化するルール一覧が表示します。各ルールは正規表現（対象URIを検索する）と条件項目から構成されます。条件項目で引数順の有無やセッションクッキ情報の排除などを指定することができます。次のセッションクッキ情報を排除することが出来ます：JSP（Javaアプリケーションサーバ）、ASP（.NET）、PHP、Broadvision（BV）。

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いのようなページが表示します：



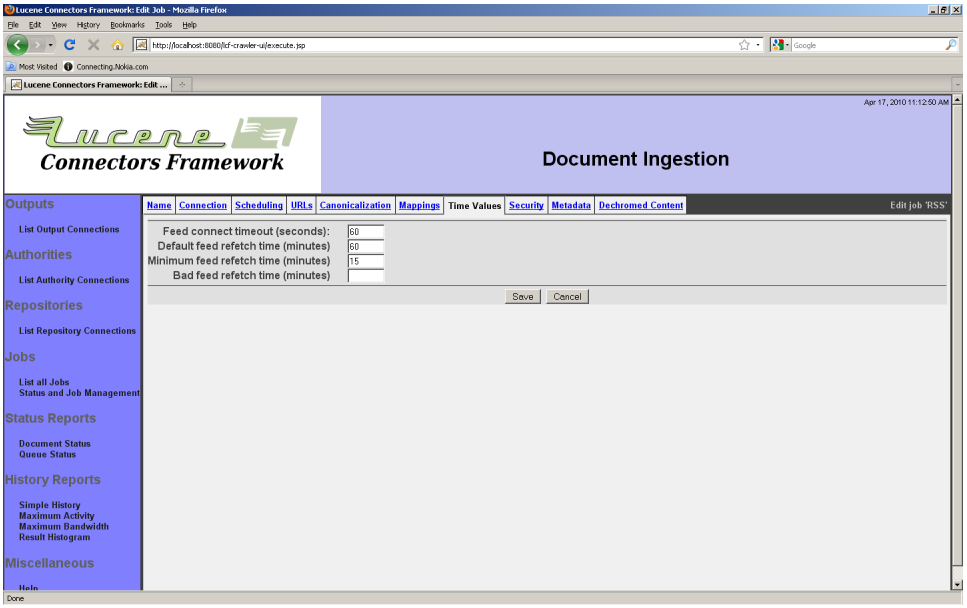
「マッピング」タブからはManifoldCFの他のタブで使われているのと同じ正規表現の仕組みで文字列の置き換えを設定することができます。マップはルールから構成されます。各ルールは一致する正規表現の式で構成されます。変換元と値は格好（「(」と「)」）で囲みます。括弧に囲まれた部分を「グループ」と言います。置き換え文字列は、固定文字と置き換えグループから構成されます。例えば、「$(1)」は最初に一致したグループを示し、「$(1l)」は最初に一致した小文字のグループを示します。同じように「$(1u)」は大文字にマップしたグループを示します。

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。

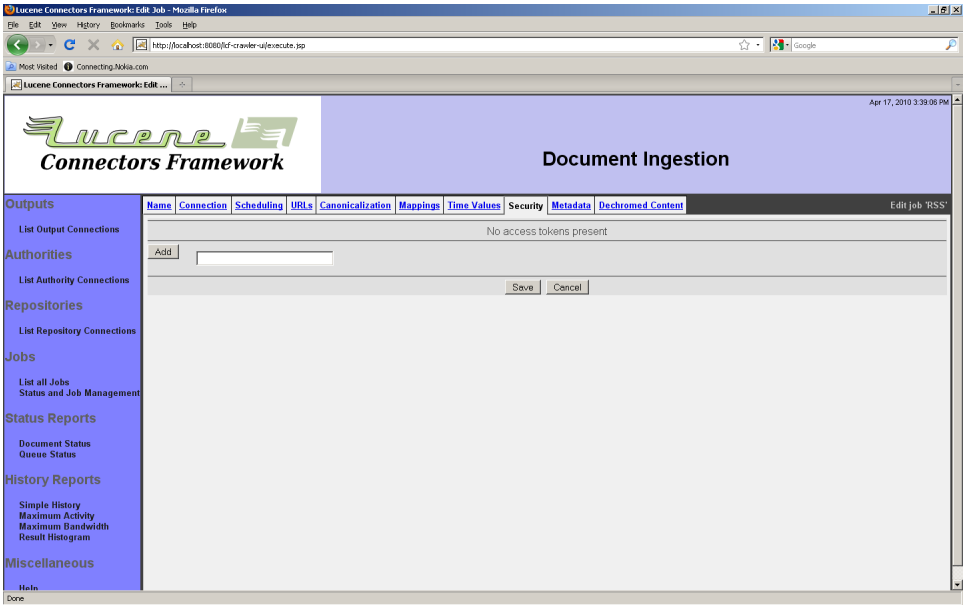
「時間」タブを選択すると次のようなページが表示します：



設定したい時間制限値を入力してください。以下んは項目の説明です：

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します：

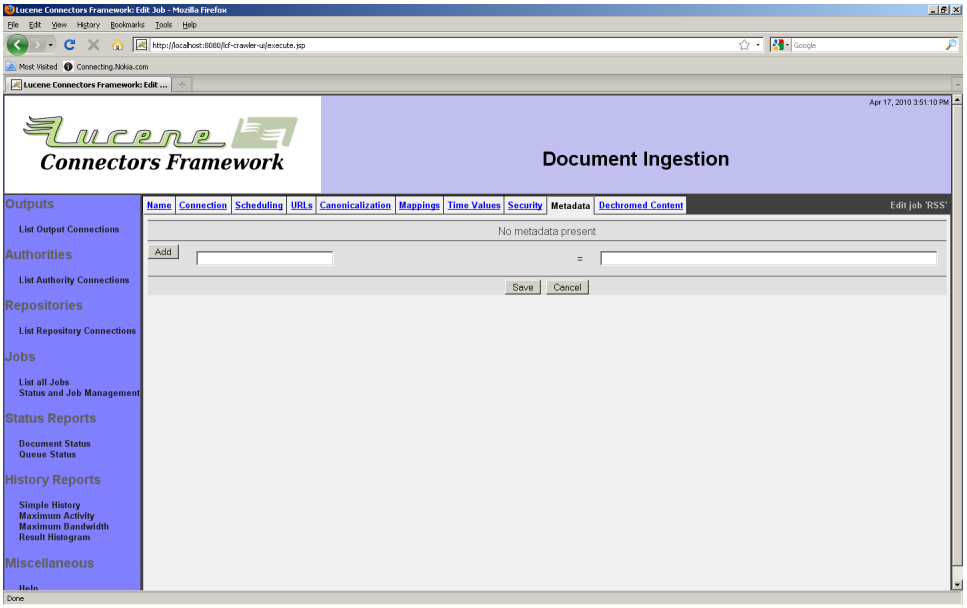


アクセストークンを追加するには、アクセルトークンの値を入力して「追加」ボタンを押下してください。アクセストークンが無い場合は、ジョブのセキュリティは無効とされます。

「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：

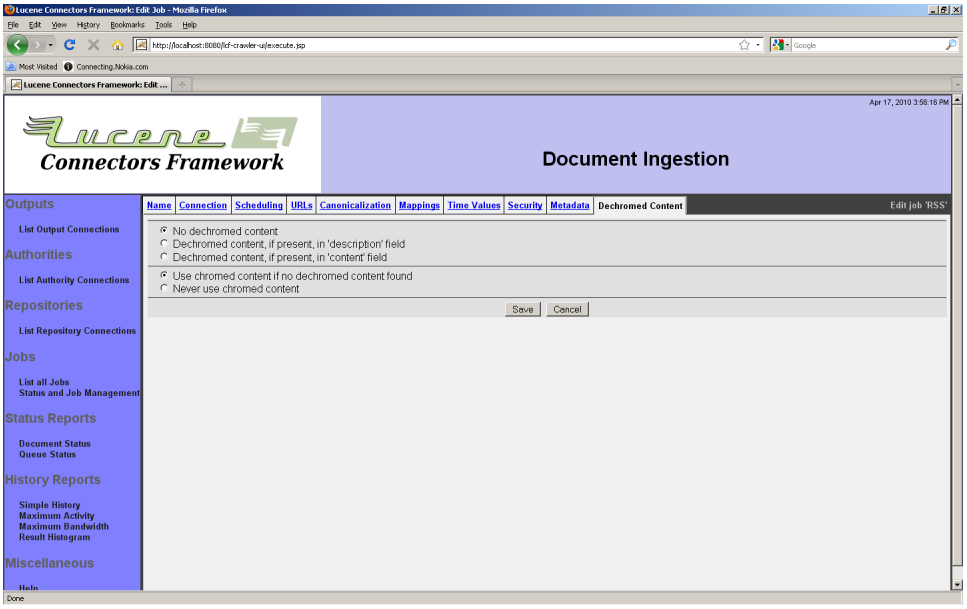
名前	説明
PubDate	コンテンツが作成された日時（1970年1月1日からのミリ秒）。作成日を取得できない場合は、コンテンツを取得した日時になります。
Source	コンテンツの名前。
Title	フィード内のコンテンツの題名。
Category	フィード内のコンテンツの分類。

「メタデータ」タブからその他に任意のメタデータを追加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：



メタデータ名と値を入力して「追加」ボタンを押下すると一覧に追加されます。

「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します：



コネクションで利用するモードを選択してください。

4.3 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

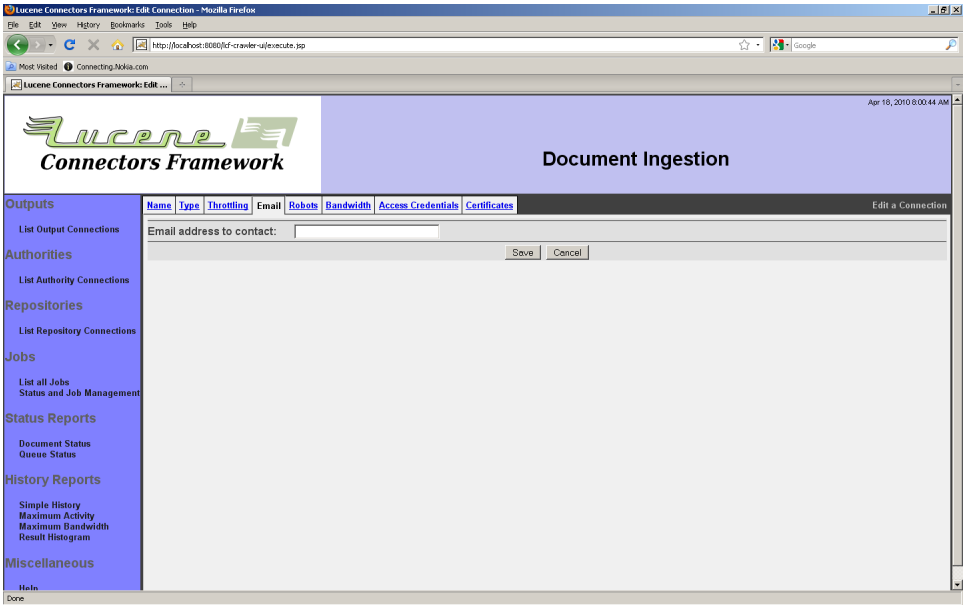
WebコネクションタイプはRSSコネクションタイプと以下の機能が異なります：

- 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－－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。

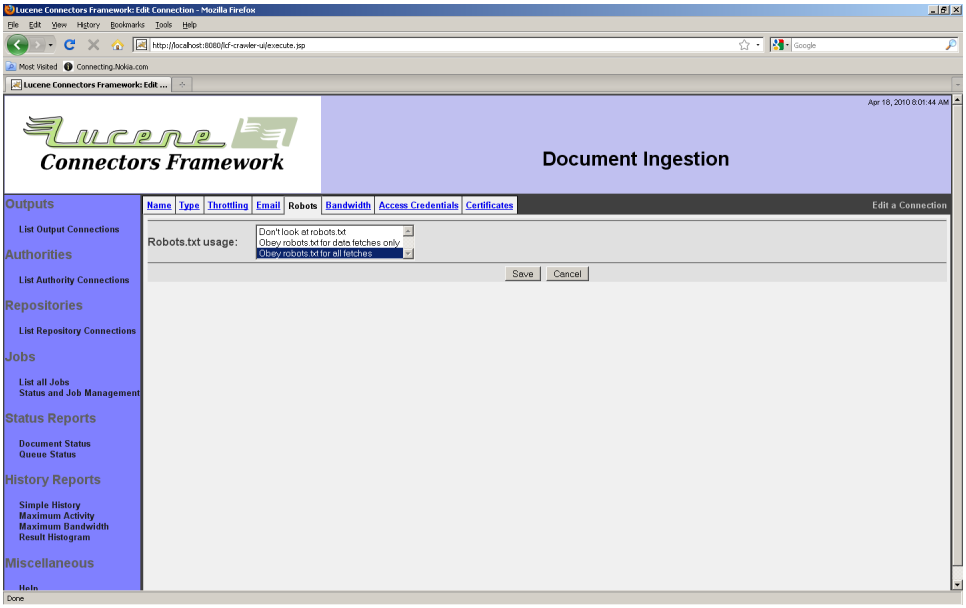
Webコネクションタイプを選択すると次のタブが表示します：「メール」、「ロボット」、「バンド幅」、「認証」、「証明書」。「メール」タブを選択すると次のようなページが表示されます：



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、ウエ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

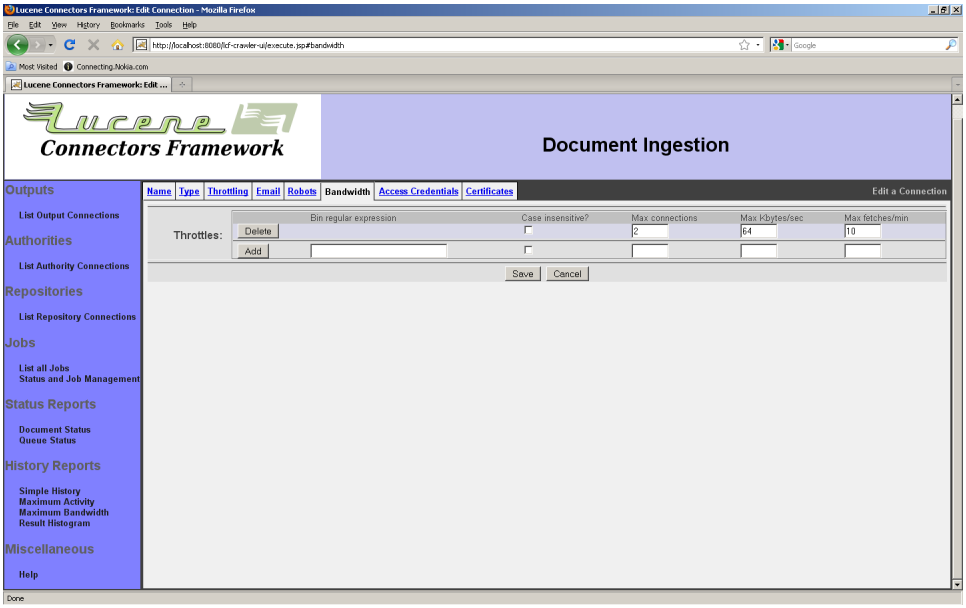
「ロボット」タブを選択すると次のようなページが表示します：



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url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します：



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。

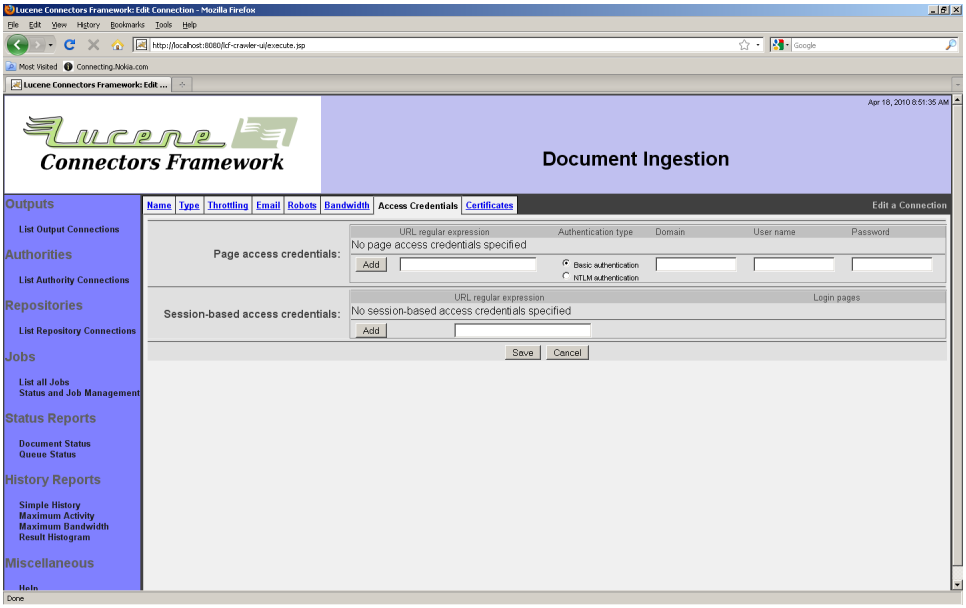
ルールを追加するには、正規表現と制限値を入力して「追加」ボタンを押下してください。

「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」タグで設定します。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しています：



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。サイト毎に次い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れです。

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又は再取得する制御はできません。

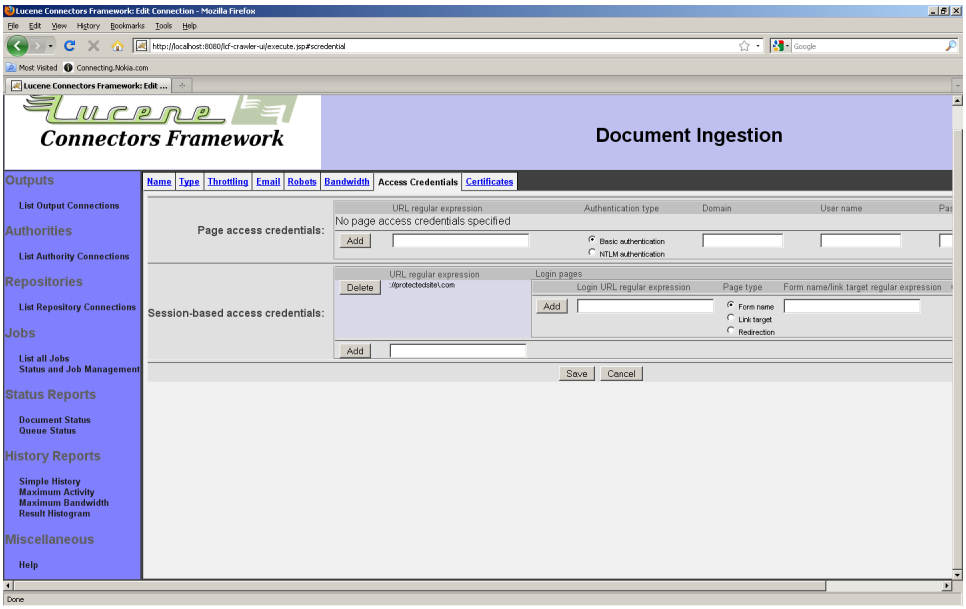
ログインページのURI及び特徴な内容からログインページを示します。例えば、セッションが無効になった場合はログインページにリダイレクトするサイトもあります。このような場合は、コンテンツを取得するよりも、ログインページへのリダイレクト情報を取得します。一般的には、ログインページ及びリダイレクト情報をコンテンツと区別して索引を作らないようにします。このような場合は、3つのログイン情報を登録します：一つはログインページへのリダイレ

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：

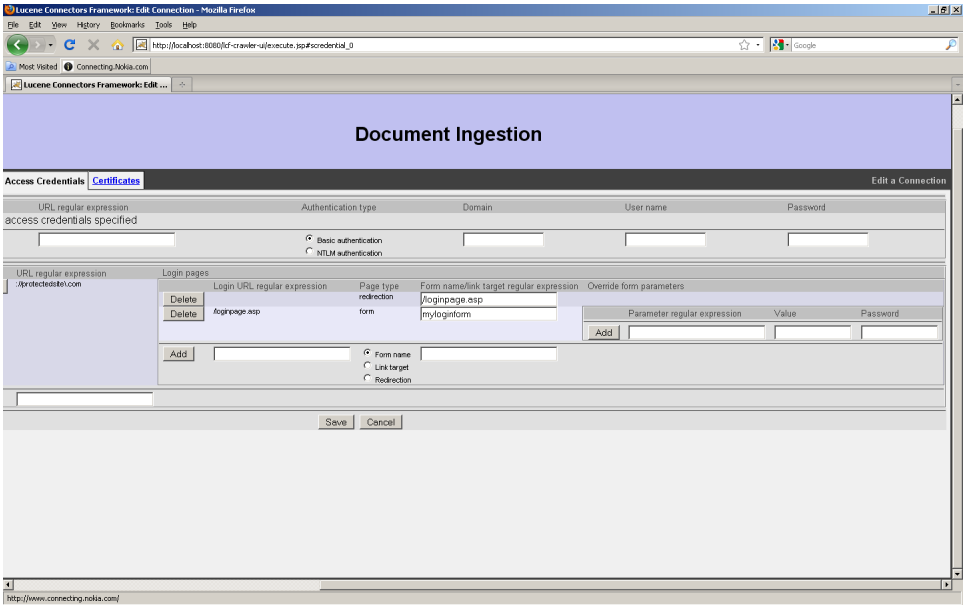
- 特定のURI（正規表現に一致）へのリダイレクト
- 指定した名前（正規表現に一致）のフォーム（FORM）が存在するページ
- 特定のページへのリンク（正規表現に一致）を含んだページ

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します：



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、ターゲットリンク又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。

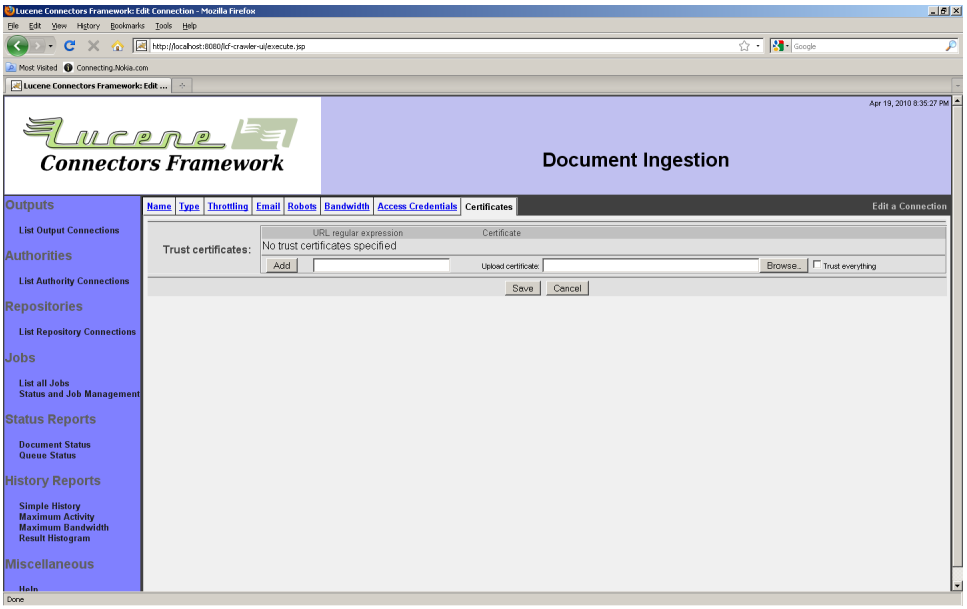
複数のログインページフォームはサイトの「ログインページ手順」です。Webコネクションは、ログインページ毎に次に取得する内容をログインページ条件で決めます。例えば、特定のURIへのリダイレクトの場合は、リダイレクトURIを取得します。フォームの場合は、フォームのactionで指定したページを取得します。ターゲットへのリンクの場合は、ターゲットURIを取得します。ログインページ手順の最後にはWebコネクションがログイン手順を開始する前に元々取得するページを取得します。

セッション認証をデバッグするい場合は、Webコネクションの簡易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。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処理が失敗又は不完了のイベントは負の値で記録されます。

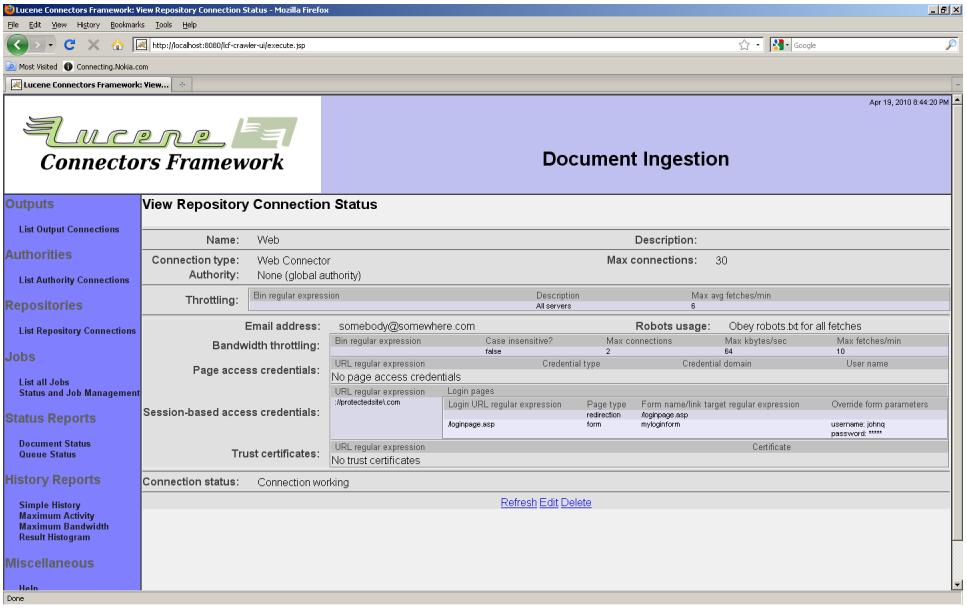
Begin login	ログイン手順を実行する場合に記録されます。ログイン手順が実行されると、ログインが完了するまでは他の保護されたサイトからコンテンツは取得されません。
End login	ログイン手順から元のコンテンツの取得に戻った時に記録されます。元のコンテンツの取得に戻った場合は、サイトから平行してコンテンツの取得を再開します。

「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します：



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。

入力しあた後に「保存」ボタンを押下しると次のような設定内容の概要ページが表示します：

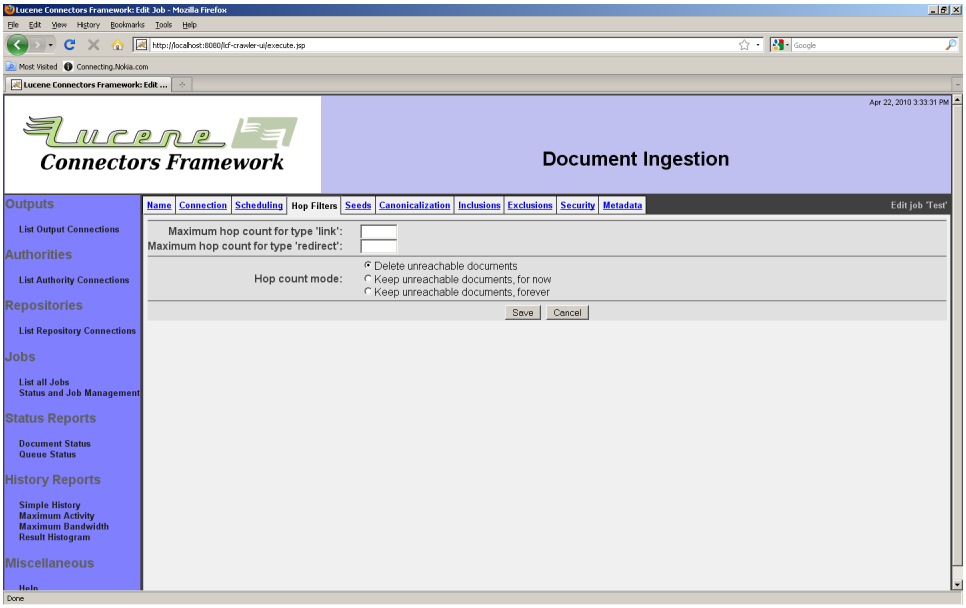


ジョブでWebタイプのリポジトリコネクションを選択した場合は、次のタブが表示します：「ホップフィルタ」、「シード」、「正規化」、「含む」、「除外」、「セキュリティ」、「メタデータ」。

「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。

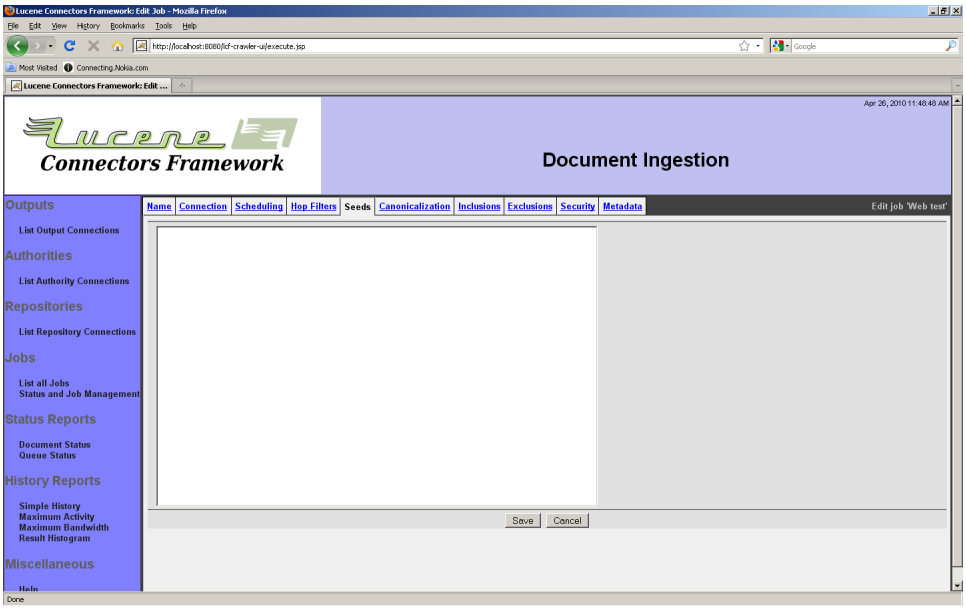
例えば、最大「リンク」ホップ数を5に設定して、「リダイレクト」ホップ数を空白にした場合は、シードコンテンツから5ホップより多いコンテンツは対象外とされます。最大「リンク」ホップ数を5に設定して、最大「リダイレクト」ホップ数を2に設定した場合は、シードコンテンツから5ホップより多くてかつリダイレクトのホップ数が2より多いコンテンツは対象外とされます。

「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します：



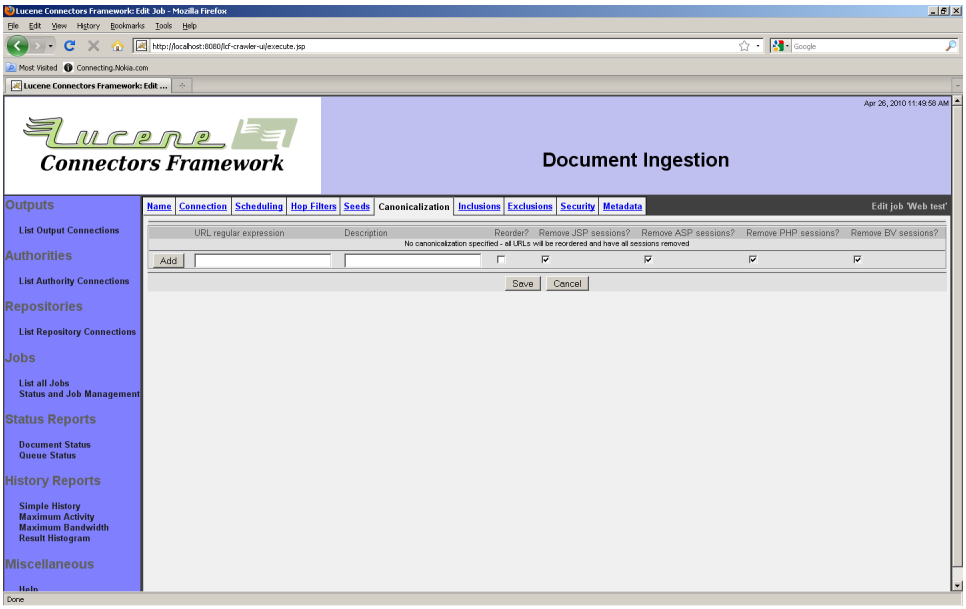
このページからルートからコンテンツのホップ数が変更された場合に行う処理を指定することができます。「読込めないコンテンツを削除」を選択すると、変更が発見されると、関係する可能性がすべてのコンテンツの深さを再計算します。再計算するとサーバに負荷が掛かります。再計算を行わないようにする場合は、一時的に行わないようにする設定と、永久に行わない設定があります。永久に行わないを選択すると、情報を削除します。

「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。「シート」タブを選択すると次のようなページが表示します：



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します：

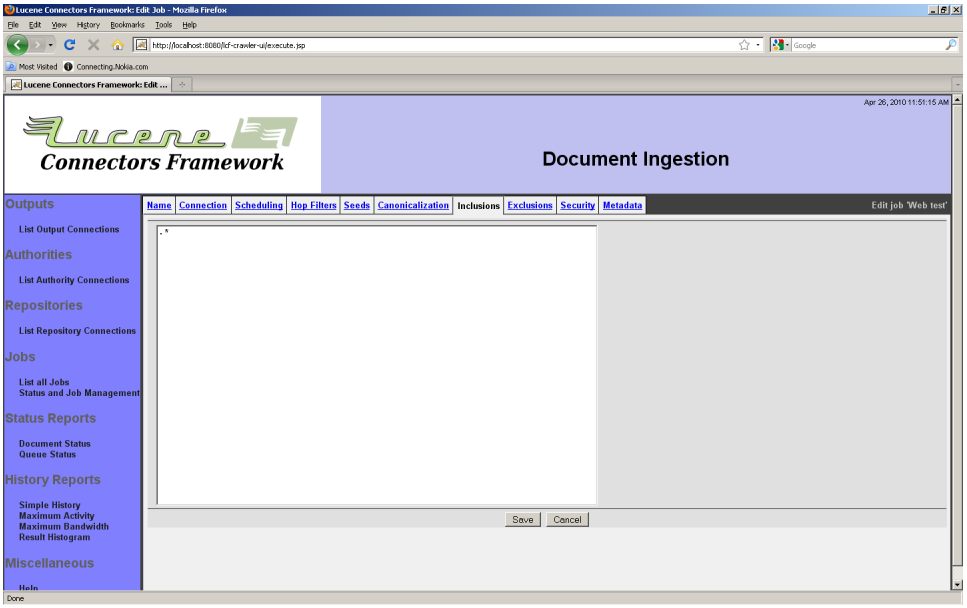


正規化するルール一覧が表示します。各ルールは正規表現（対象URIを検索する）と条件項目から構成されます。条件項目で引数順の有無やセッションクッキ情報の排除などを指定することができます。次のセッションクッキ情報を排除することが出来ます：JSP（Javaアプリケーションサーバ）、ASP（.NET）、PHP、Broadvision（BV）。

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。

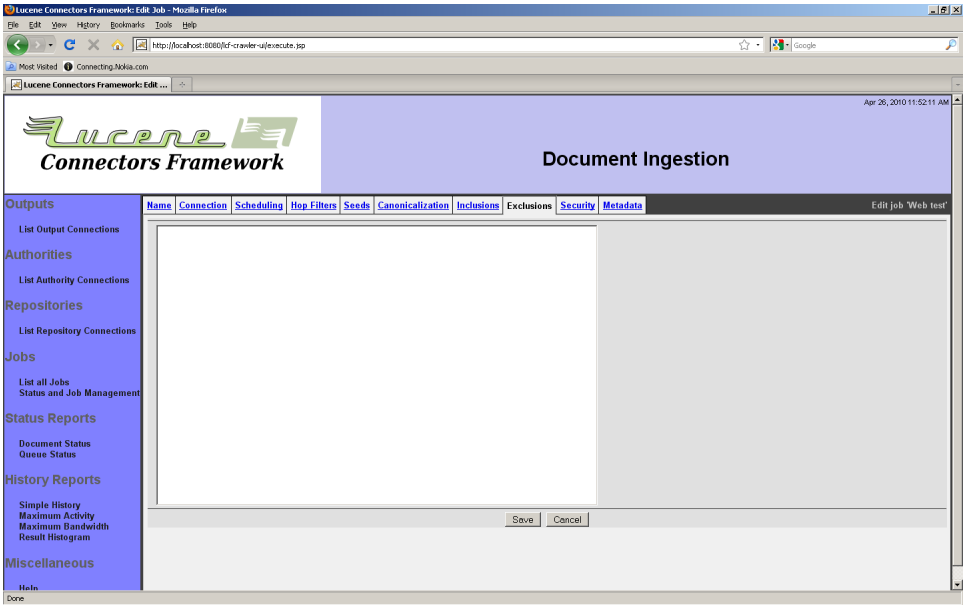
「含む」タブからWebジョブに含むURI正規表現一覧を指定することができます。「含む」タブを選択すると次のようなページが表示します：



改行区切りで0以上の正規表現を指定してください。

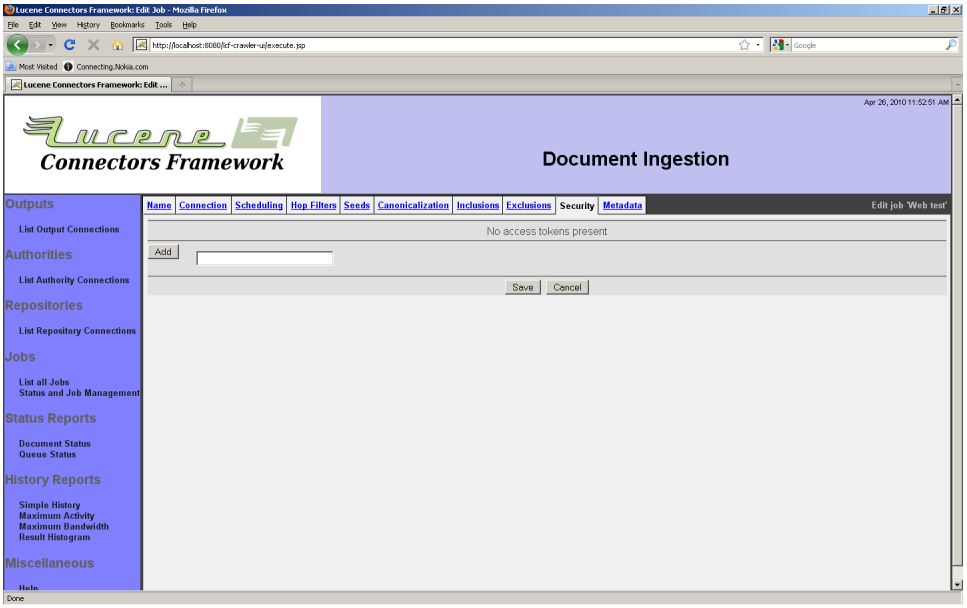
Webジョブはデフォルトでは、シードにリンクされているインターネット上のすべてのコンテンツを含みます。

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します：



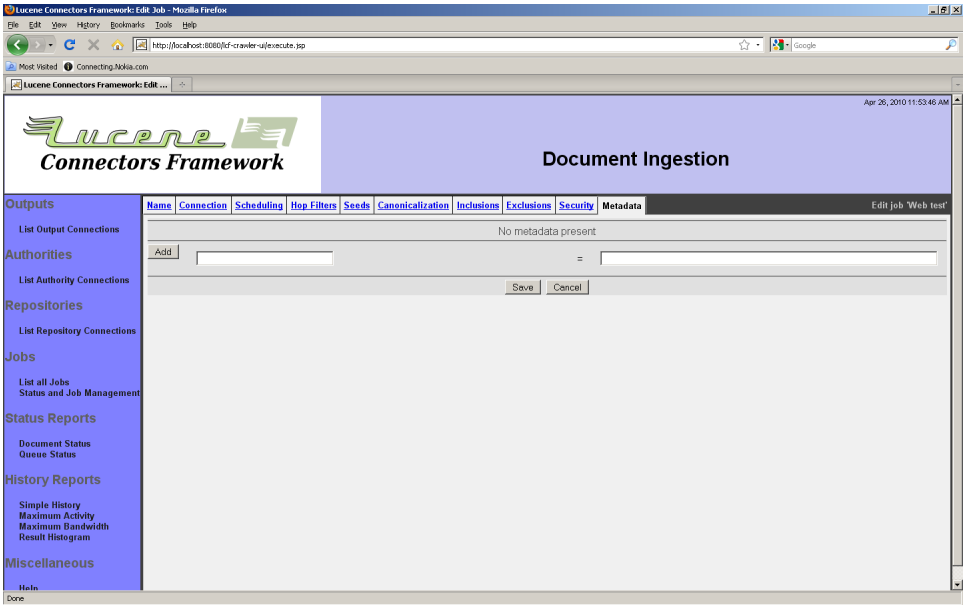
改行区切りで0以上の正規表現を指定してください。索引を作成できないコンテンツは、索引をする必要がないコンテンツを除外することを推奨します。例えば動画や映像などからは索引を作成できないので除外の対象です。

「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します：



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」タグ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します：



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。

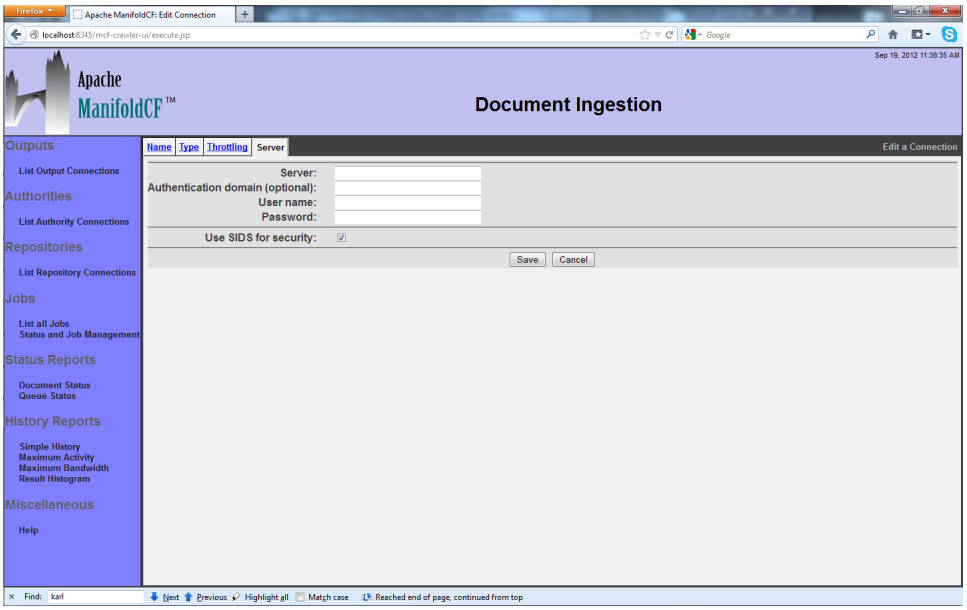
4.4 Windows Share/DFSリポジトリコネクション

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のシステムからも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又は「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に変換すること式はありません。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、事前にアクティブディレクトリ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します:

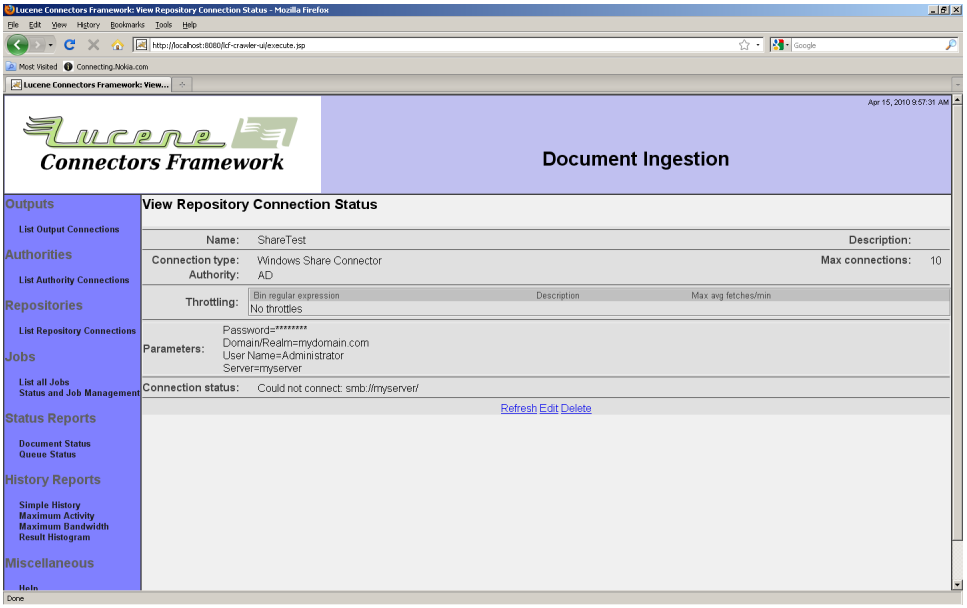


接続するサーバ名を入力してください。サーバ名は、実サーバ名又はWindowsドメインDFSルートに接続する場合はドメイン名で指定することができます。実サーバ名を入力する場合は、サーバ名に未修飾名を入力して、「ドメイン名」に完全修飾ドメイン名を入力してください。ユーザ名は未修飾名を入力して下さい(例:「Administrator@mydomain.com」ではなく、「Administrator」と入力してください)。「ドメイン名」を空白にして、「サーバ」に完全修飾サーバ名を入力する場合もあります。ただし、ドメイン名と完全修飾サーバ名の両方を入力しないでください。

"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。

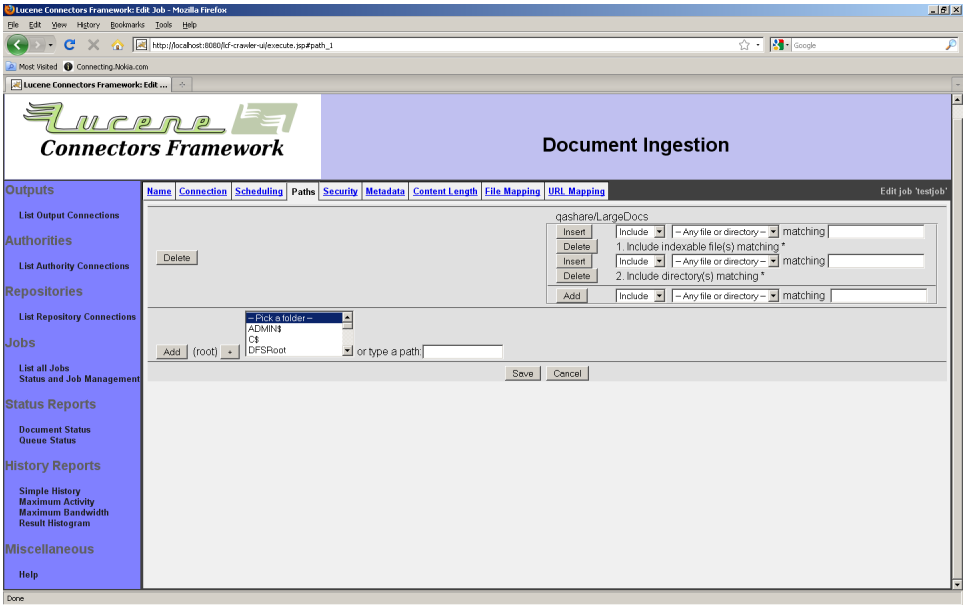
入力した後に「保存」ボタンを押下してください。次のような設定概要ページが表示します:



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します:「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します:



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。

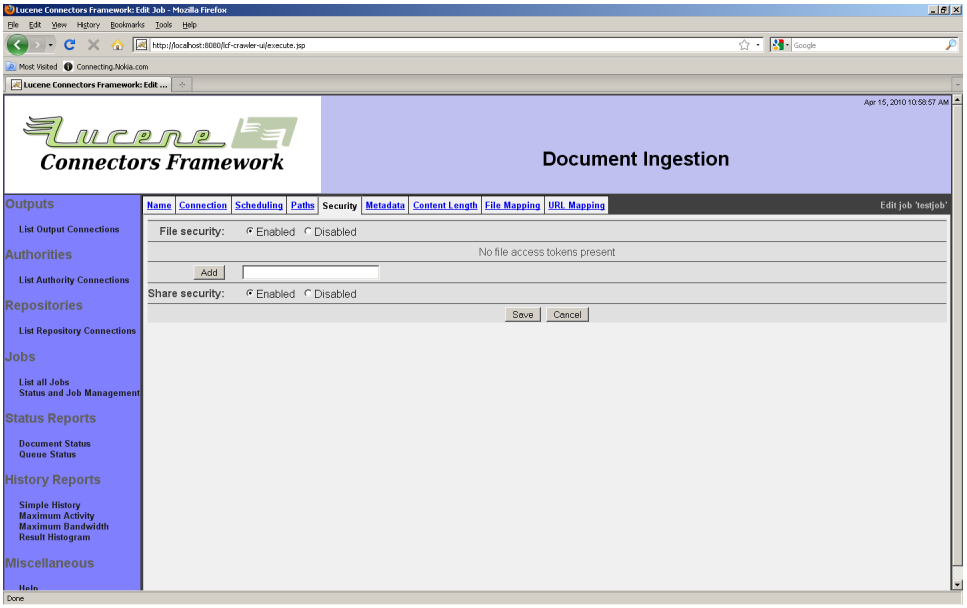
コネクションの状態が「起動」になっていることを確認してください。エラーメッセージが表示している場合は、パスは作成されません。

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」）、ファイル又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します：



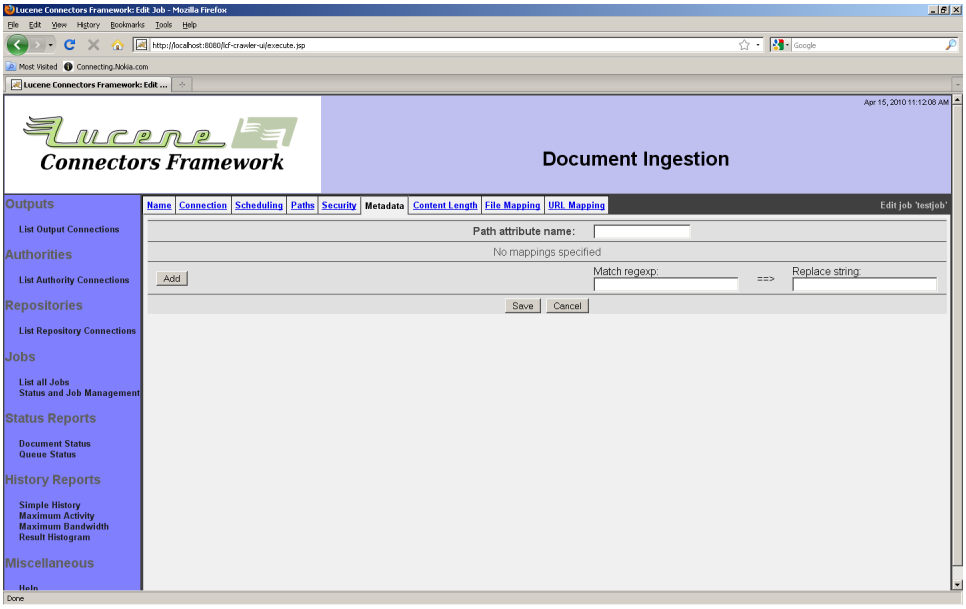
「セキュリティ」タブからは次の3機能を設定することができます：ファイルセキュリティ、共有セキュリティ、ジョブで索引が作成されたすべてのコンテンツのセキュリティトークン（セキュリティが無効の場合）。

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。殆どの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され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共有セキュリティを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します：

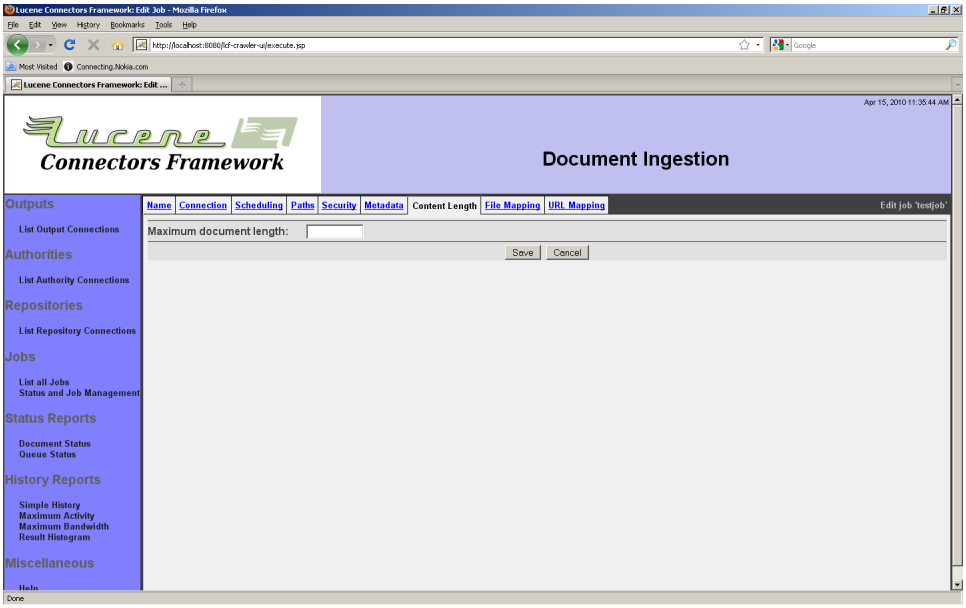


正規表現により変換されたコンテンツパスをコンテンツのメタデータとして取得することをできるようにする設定を行えます。「パス属性名」にメタデータ名を入力した後にルール一覧にルールを追加してください。各ルールは一致する正規表現の式で構成されます。変換元と値は格好（「(」と「)」）で囲みます。括弧に囲まれた部分を「グループ」と言います。置き換え文字列は、固定文字と置き換えグループから構成されます。例えば、「$(1)」は最初に一致したグループを示し、「$(1l)」は最初に一致した小文字のグループを示します。同じように「$(1u)」は大文字にマップしたグループを示します。

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。

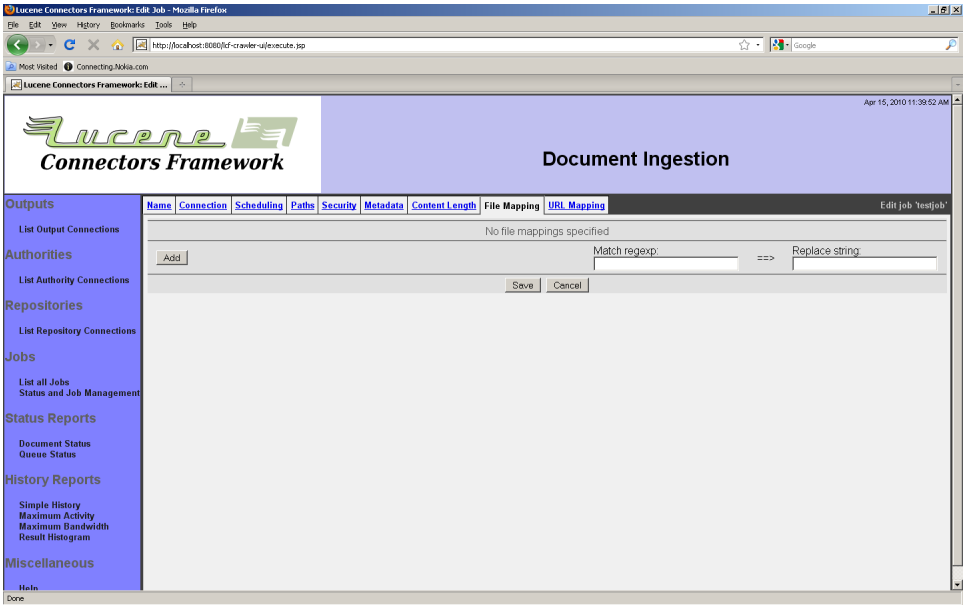
1つ以上のルールが存在する場合は、上から実行され、上のルールの結果は下のルールで変更されます。

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



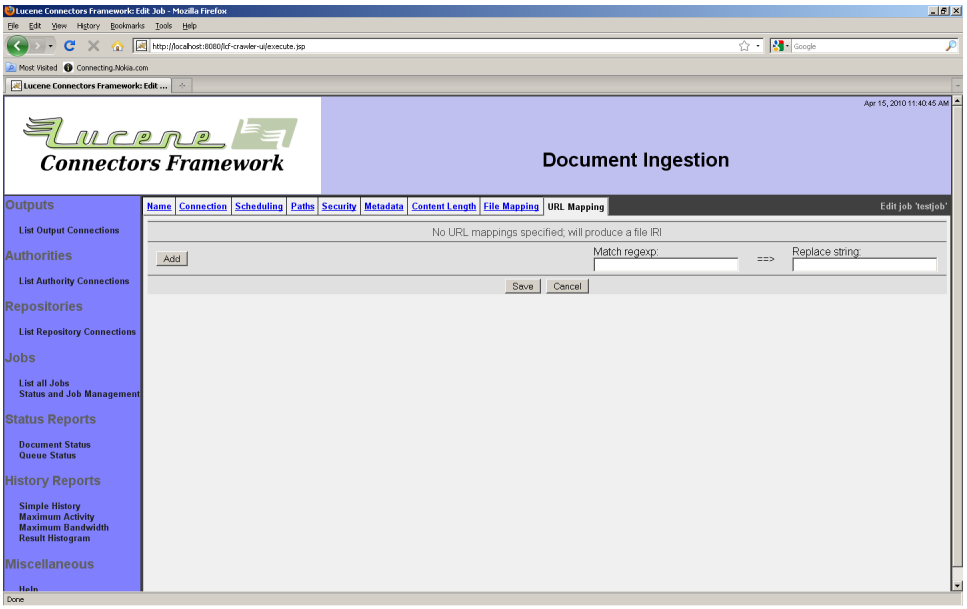
ジョブが長いコンテンツの作成を作成しないように、コンテンツを処理する最大長さを指定することができます。指定した長さより長いコンテンツは、指定した長さで切り捨てられます。最大長さを入力してください。最大長を設定しない場合は、空白にしてください。

「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いのようなページが表示します：

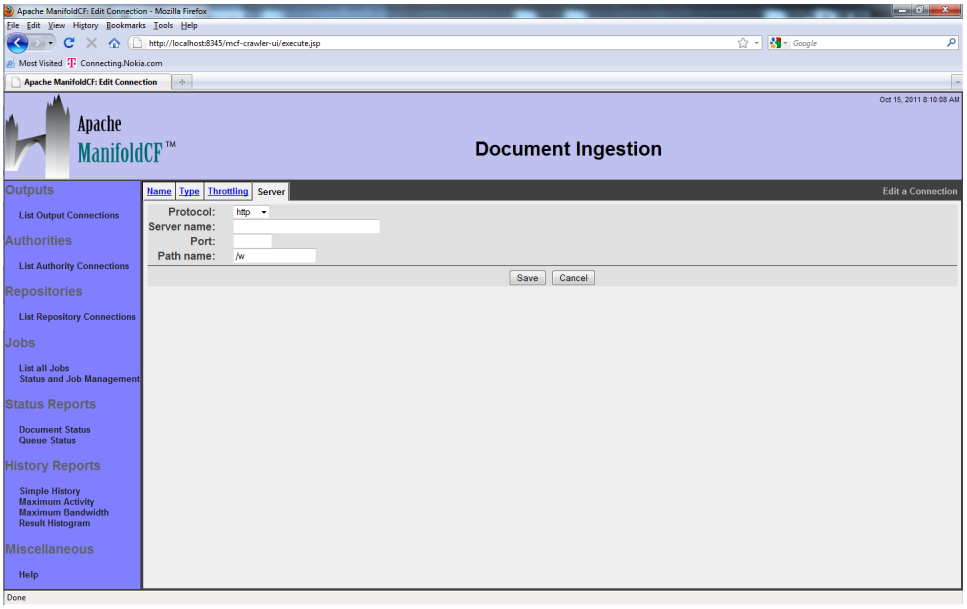


URLマップはパス属性マップと同じように設定します。マップが存在しない場合は、ファイルパスは標準ファイルIRIに変換されます。マップが存在する場合は、Windows共有HTTPサーバを介して取得できる正しいURIへ変換されることを前提にします。

4.5 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。

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」タグが表示します。「サーバ」タブを選択すると次のようなページが表示します：



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」に入力して、ポート番号を項目「ポート」にください。最後に、WikiのURIを項目「パス名」に入力してください。URI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。

4.6 汎用データベースリポジトリコネクション

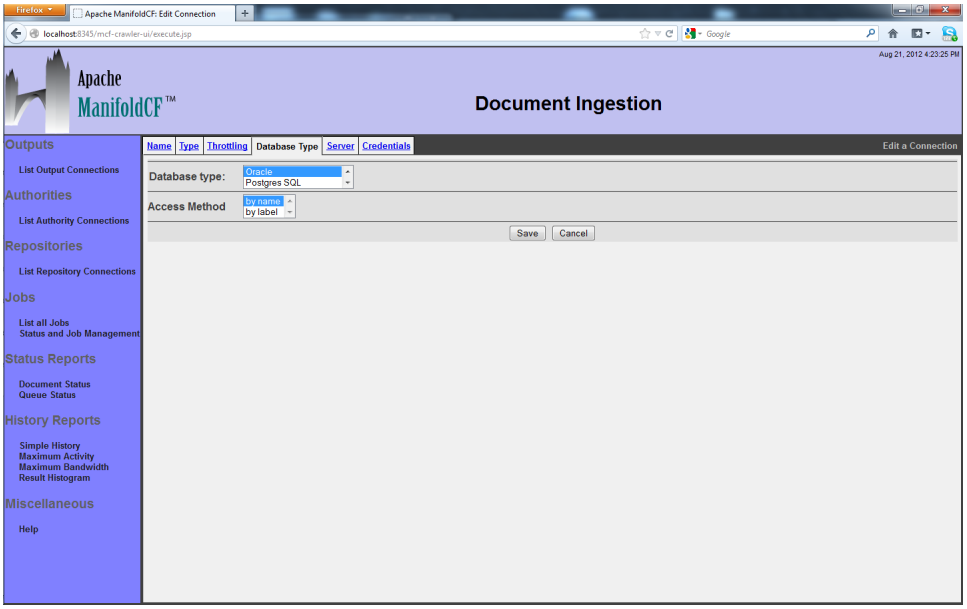
汎用データベースコネクションタイプは次のデータベースのテーブルの内容から索引を作成します:

- 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エ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。

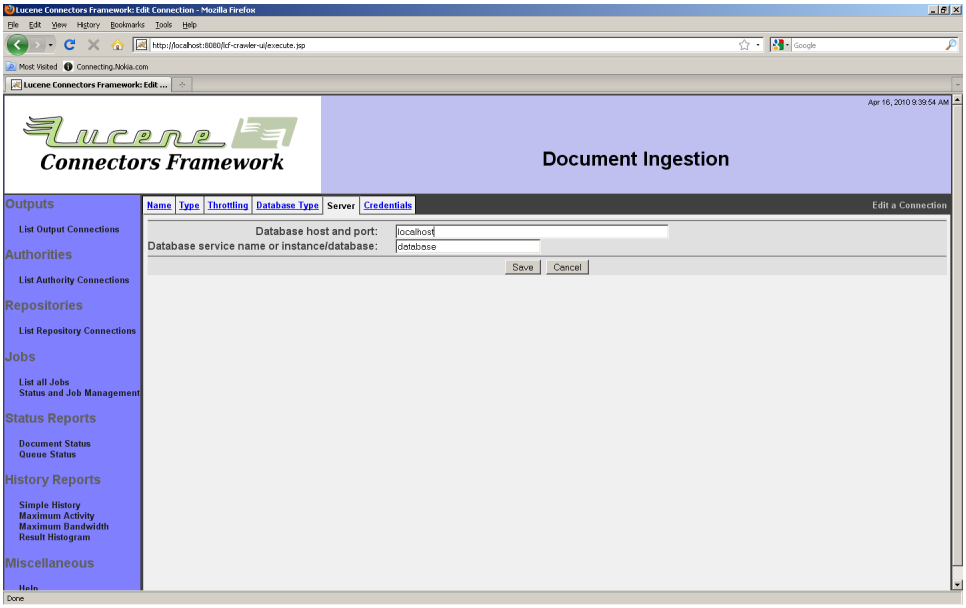
リポジトリコネクション編集ページから汎用データベースコネクションを選択すると3つのタブが表示します：「データベースタイプ」、「サーバ」、「認証」。「データベースタイプ」タブを選択すると次のようなページが表示します：



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。

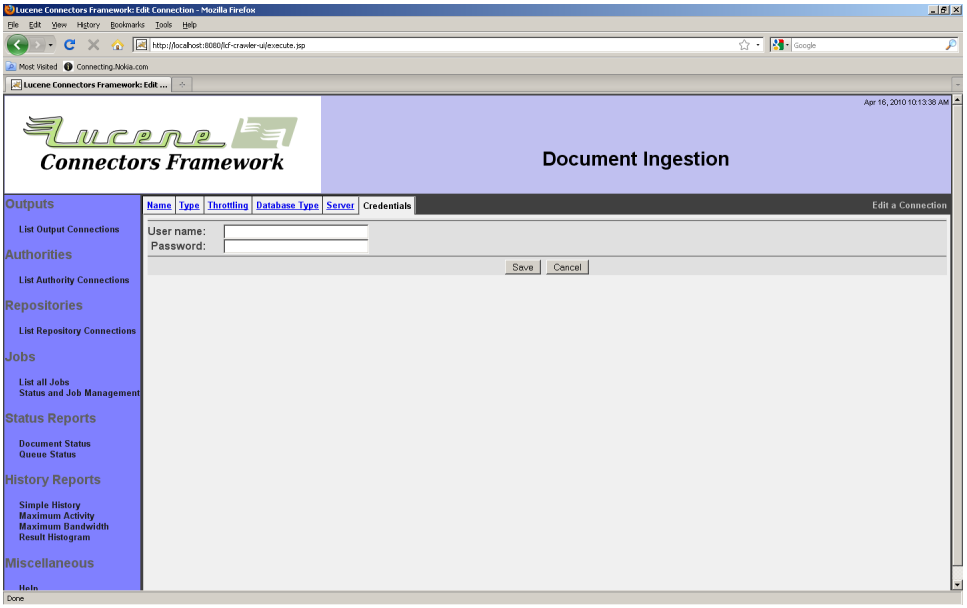
「サーバ」タブを選択すると次のようなページが表示します：



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。

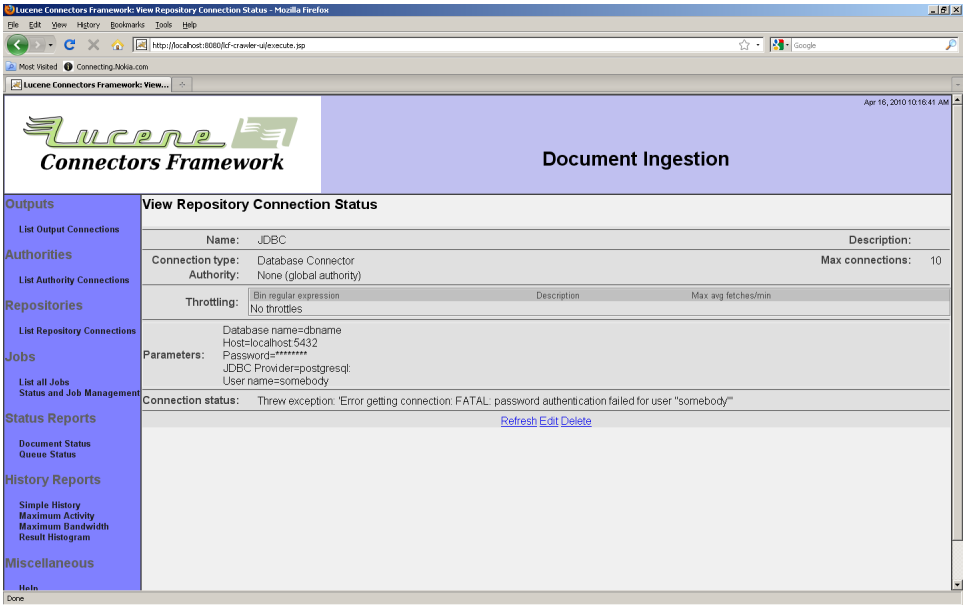
サービス名及びインスタンス名には利用するデータベースのインスタンス名を入力してください。Oracle及びPostgresqlの場合は、データベース名を入力してください。SQL Serverの場合は「my-instance-name/my-database-name」のように入力してください。SQL Serverのデフォルトのインスタンスを利用する場合は、データベースのみを入力してください。

「認証」タブを選択すると次のようなページが表示します：



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。

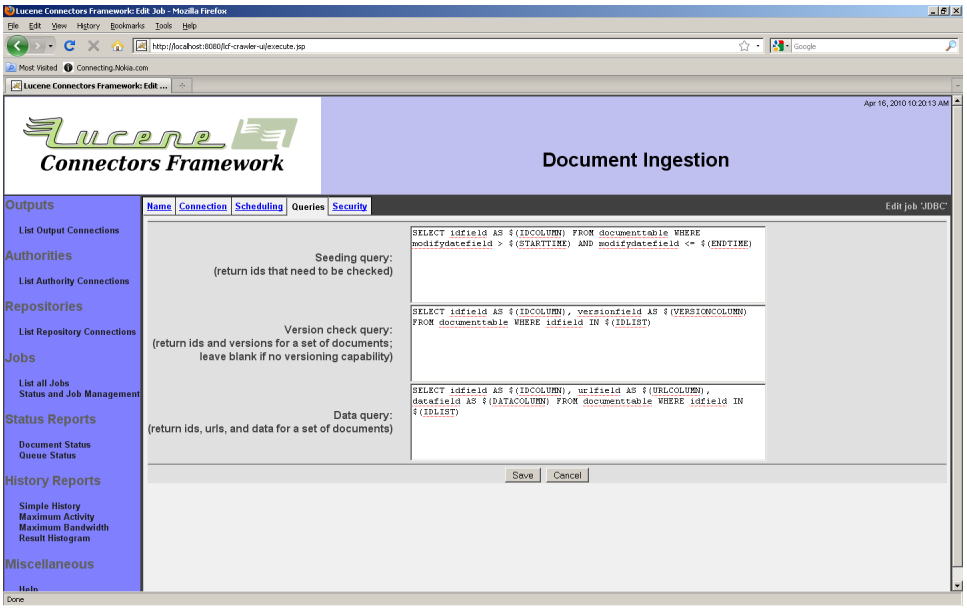
入力した後に「保存」ボタンを押下してください。次のような設定内容概要ページが表示します：



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」タブが表示します。

「クエリー」タブを選択すると次のようなページが表示します：



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てい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（下の参照）。
- 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。コンテンツ情報には、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。

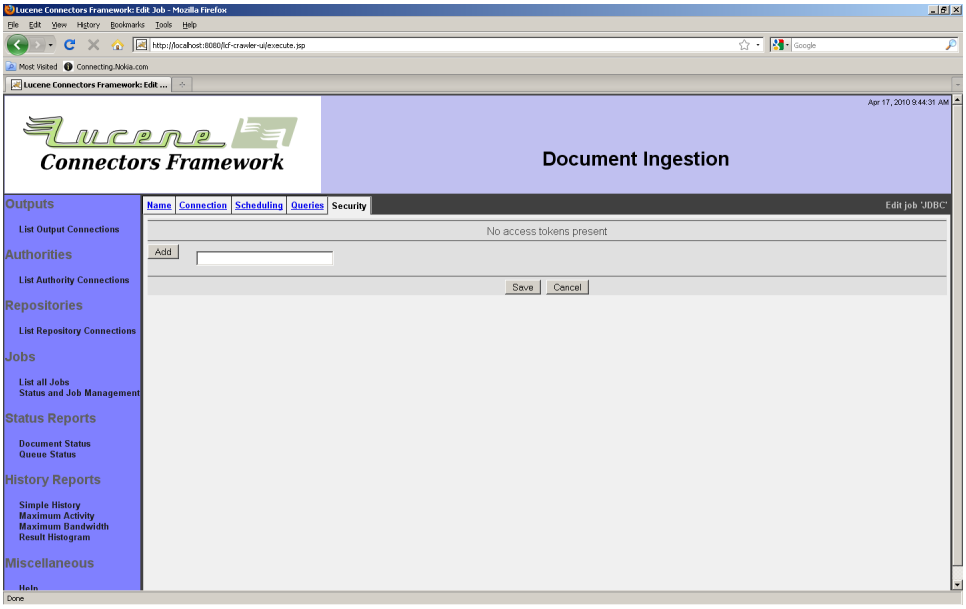
以下は「$(STARTTIME)」と「$(ENDTIME)」を他形式の日付と時間に変換するサンプルクエリーの部分です。先頭の列はクエリー句が対応するSQLデータベースです。2列目はクエリーの出力データ型です。3列目は、「$(STARTTIME)」を利用した例です。これらはクエリーの例です。変更したデータ型が2列目のデータ型と異なる場合は、日付比較は正しくされない場合もあります。

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します：



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。

4.7 IBM FileNet P8リポジトリコネクション

4.8 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

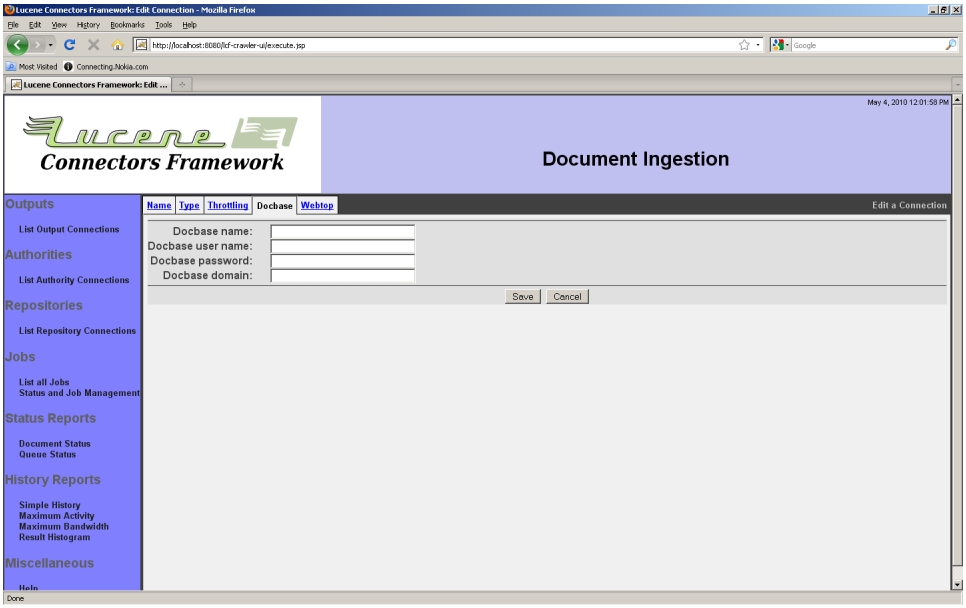
Serverインスタンスのすべてのコンテンツを取得することができます。複数のContent Serverインスタンスのコンテンツの索引を作成する場合は、サーバインスタンス毎にコネクションを定義する必要があります。

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。

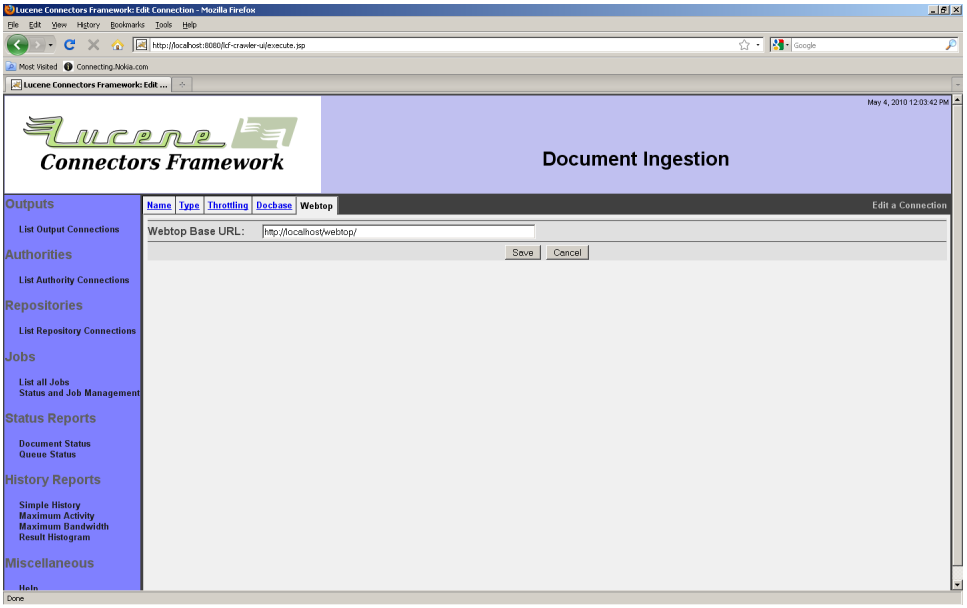
Documentumコネクションを選択すると次のタブが表示します:「Docbase」、「Webtop」。「Docbase」タブからは接続するコンテンツサーバの指定と、認証情報の設定ができます。索引を作成した後に、「Webtop」タブからはコンテンツサーバの内容を表示するWebtopサーバの指定することがでけいます。

「Docbase」タブを選択すると次のようなページが表示します:



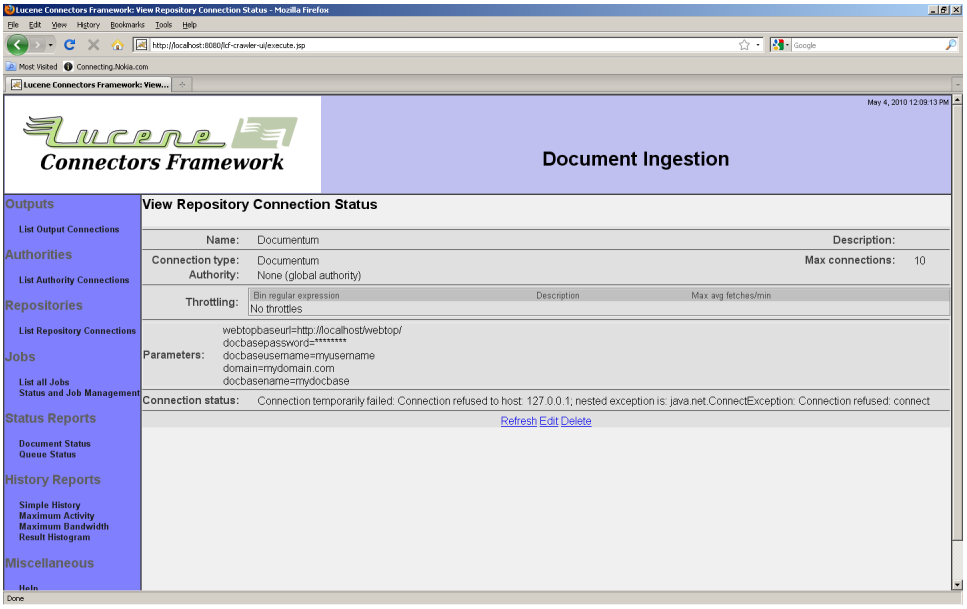
コンテンツサーバDocbaseインスタンス名と認証情報を入力してください。コンテンツサーバインスタンスがアクティブディレクトリと統合されていない場合は、「ドメイン」を空白にしてください。

「Webtop」タブを選択すると次のようなページが表示します;



Webtopインスタンスのベース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込みの対象にします。

「ドキュメントタイプ」タブからは、読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。

「コンテントタイプ」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。変換元と値は格好(「(」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。

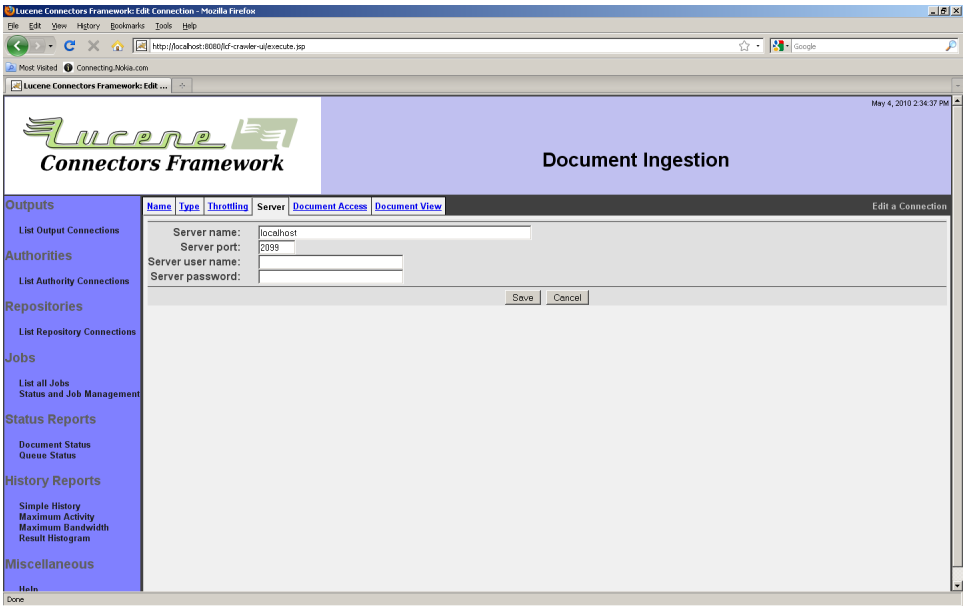
4.9 OpenText LiveLinkリポジトリコネクション

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。

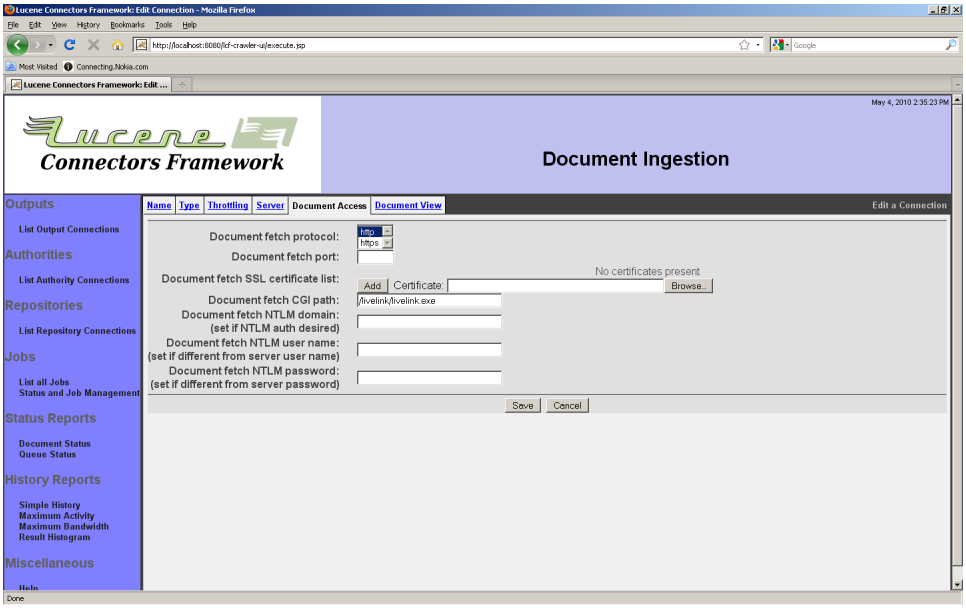
LiveLinkコネクションを選択すると次のタブが表示します:「サーバ」、「ドキュメントアクセス」、「ドキュメント表示」。「サーバ」タブからは、接続するLiveLinkサーバの選択と接続するための認証情報を設定することができます。「ドキュメントアクセス」タブからはLiveLinkからコンテンツを取得するためのLiveLinkのwebインタフェースの情報を設定します。「ドキュメント表示」タブからは、検索結果を表示する取得したコンテンツのURIを作成方法を指定します。

「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:



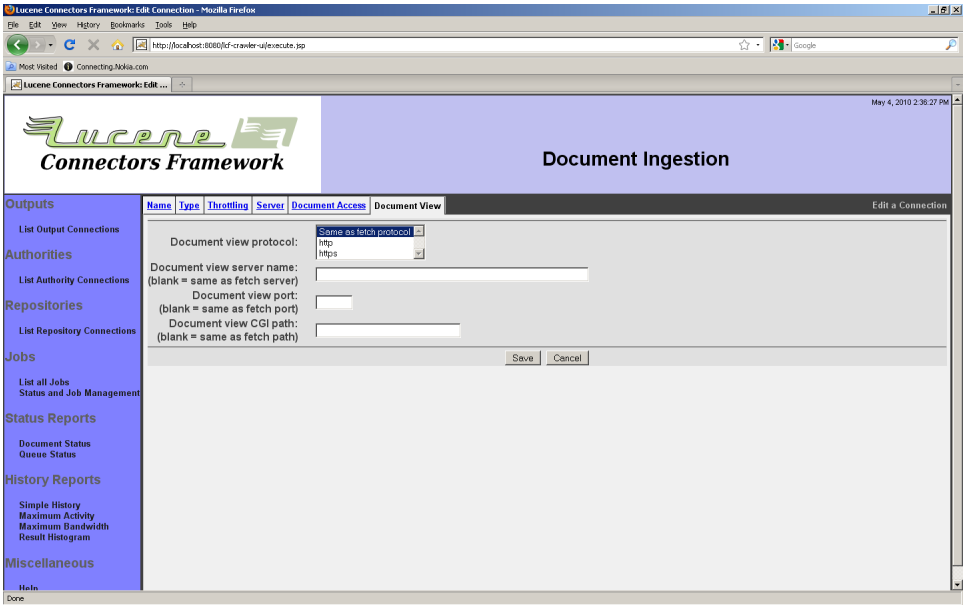
LiveLinkサーバ名、ポート番号、認証情報を入力してください。

「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します:



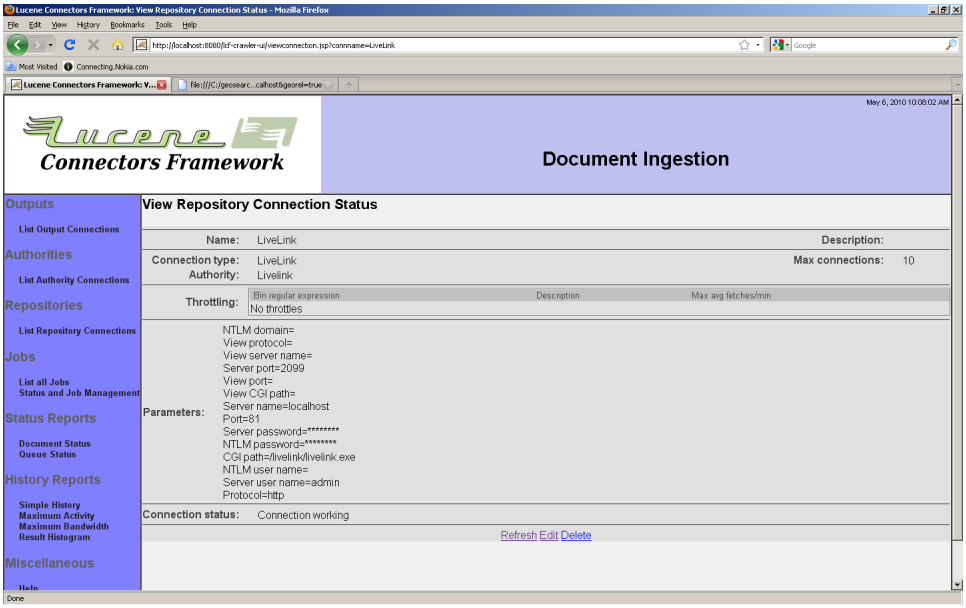
サーバ名は「サーバ」タブに入力した値と同じと想定されます。プロトコルを選択してください。LiveLinkサーバはHTTPデフォルト以外のポート番号を利用している場合はポート番号を入力してください。LiveLinkサーバがNTLM認証を利用している場合は、アクティブディレクトリのユーザ名、パスワードとドメイン名を入力してください。LiveLinkサーバがHTTPSを利用している場合は、「追加」ボタンを押下して証明書をアップロードしてください。（サーバの証明書を使うこともできますが、サーバの証明書は変わる可能性がありますので注意してください。）

「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します：



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。

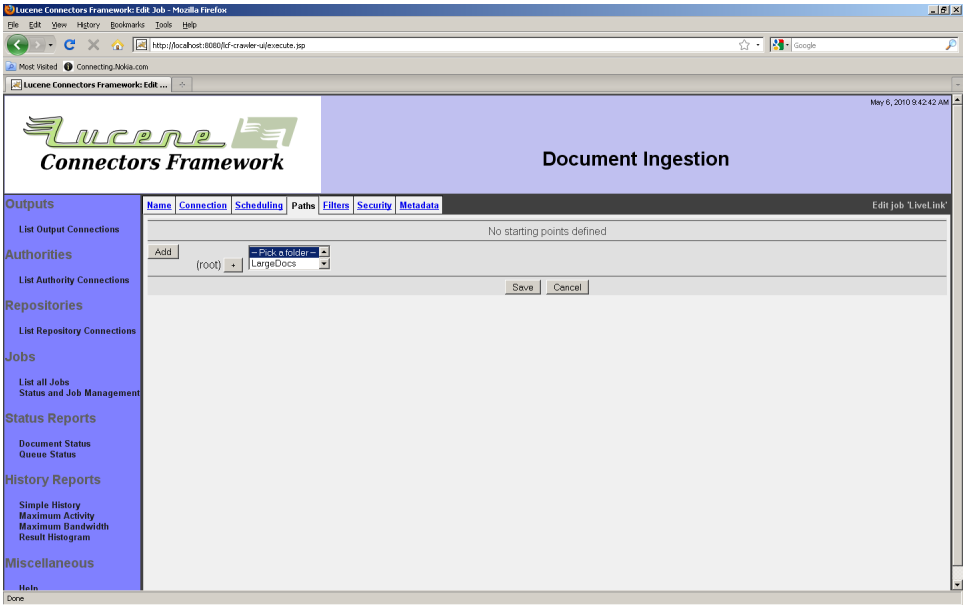
入力した後に「保存」ボタンを押下してくだしあ。次のような設定内容概要ページが表示します：



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。

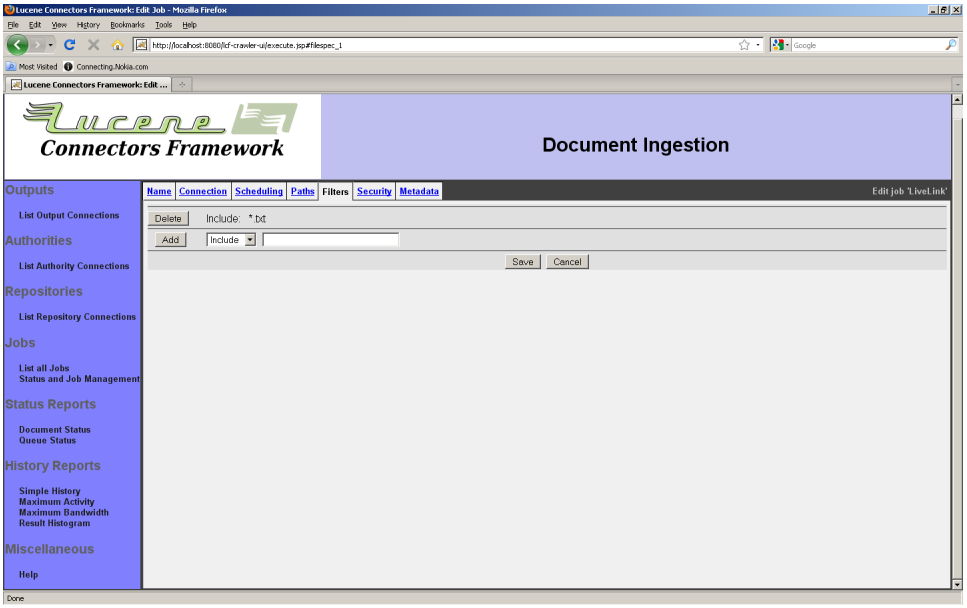
ジョブでLiveLinkコネクションを選択すると次のタブが表示します：「パス」、「フィルタ」、「セキュリティ」、「メタデータ」。

「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します：



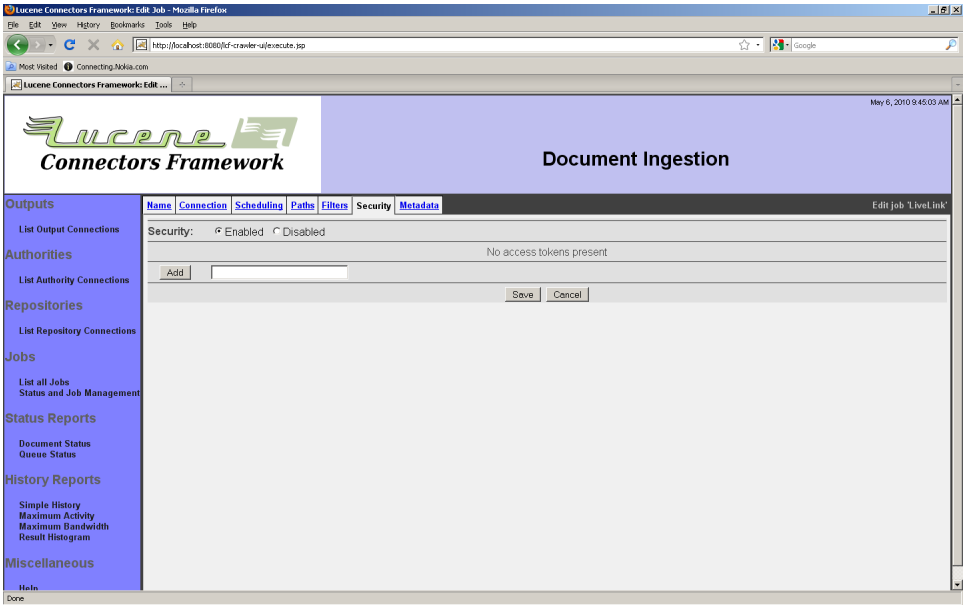
ドロップダウンリストから選択して「+」ボタンを押下してパスを作成してください。パスを作成し終わったら「追加」ボタンを押下して一覧に追加してください。

「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」）、ファイル又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。



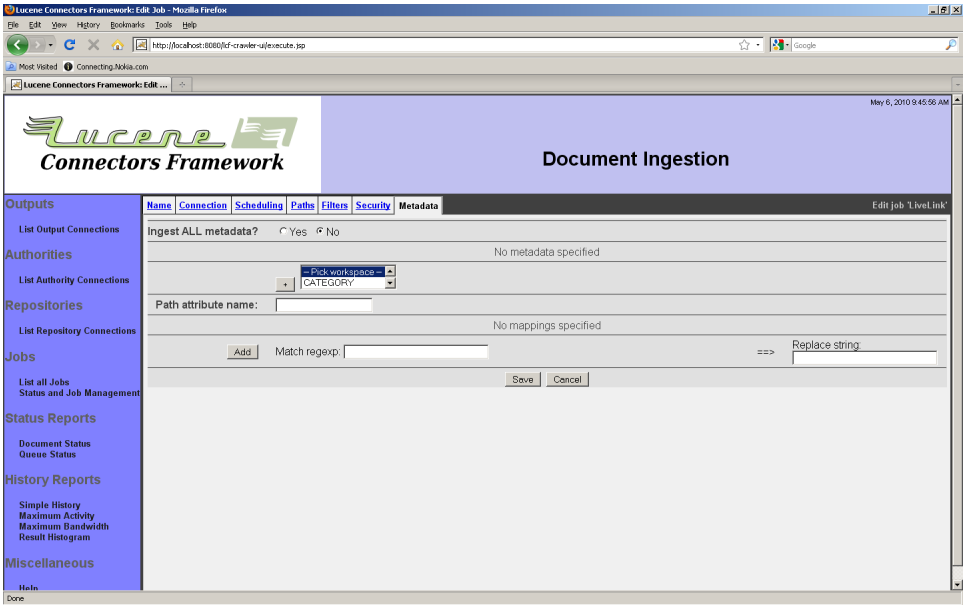
一致させる値を入力し、一致した場合の処理を選択した後に「追加」ボタンを押下してルールをフィルター覧に追加してください。

「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します：



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します：



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ー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しあ。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。変換元と値は格好（「(」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。

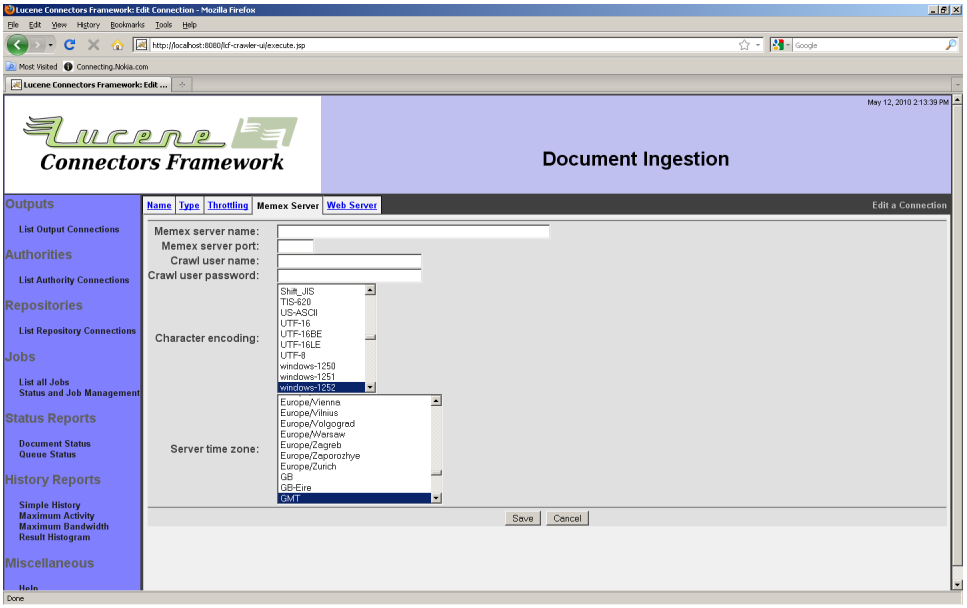
1つ以上のルールが存在する場合は、上から実行され、上のルールの結果は下のルールで変更されます。

4.10 Memex Patriarchリポジトリコネクション

Memex Patriach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。

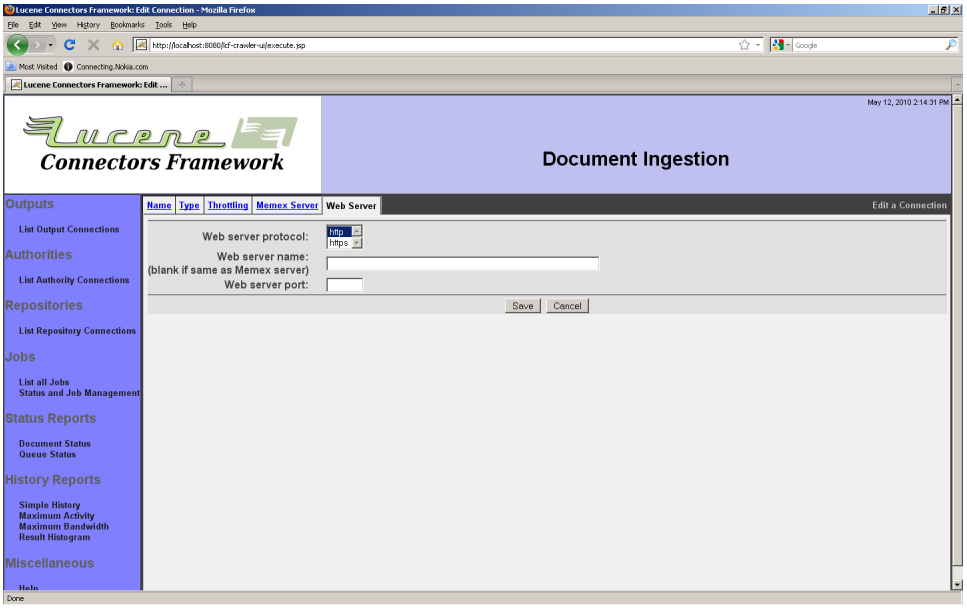
Memexコネクションで処理するコンテンツのセキュリティはMemex権限で管理されています。Memex権限コネクションの設定については「Memex Patriarch権限コネクション」を参照してください。

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い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:



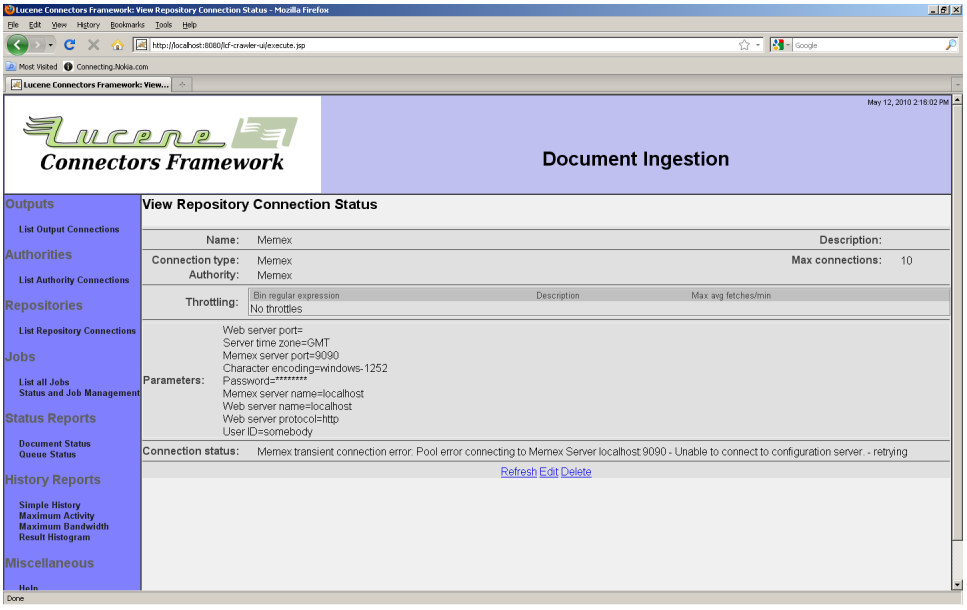
Memexサーバ名、コネクションポート番号、コンテンツを取得できる権限をもつMemexユーザの認証情報を入力して下さい。また、Memexサーバで利用している文字エンコーディングとタイムゾーンも入力してください。

「Webサーバ」タブを選択すると次のようなページが表示します:



Memexコンテンツ毎に一意性の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：



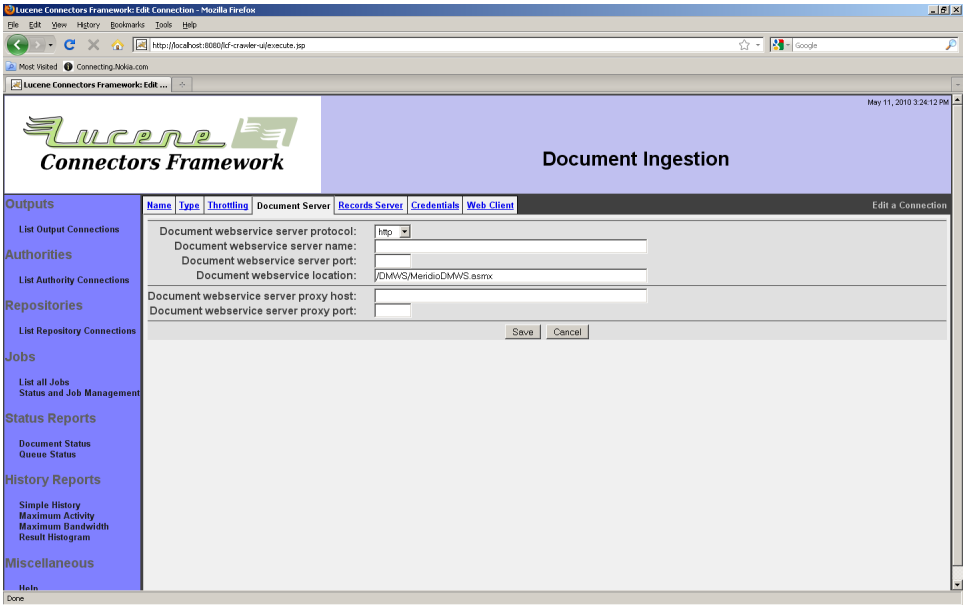
ジョブからMemexコネクションを選択すると次のタブが表示します：「レコード条件」、「エンティティ」、「セキュリティ」。

4.11 Meridioリポジトリコネクション

Autonomy Meridioコネクションは、Merido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動させ、レコードサービスを別のサーバで稼動させる）。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。

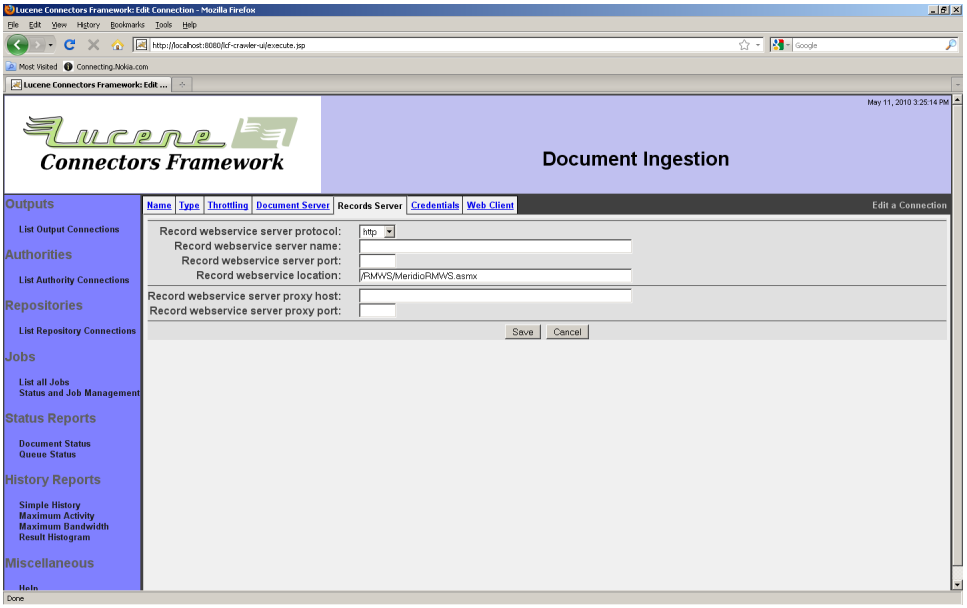
リポジトリコネクションページからMeridioコネクションを選択すると次のタブが表示します：「ドキュメントサーバ」、「レコードサーバ」、「Webクライアント」、「認証」。「ドキュメントサーバ」タブを選択すると次のようなページが表示します：



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。プロキシを利用されている場合は、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。

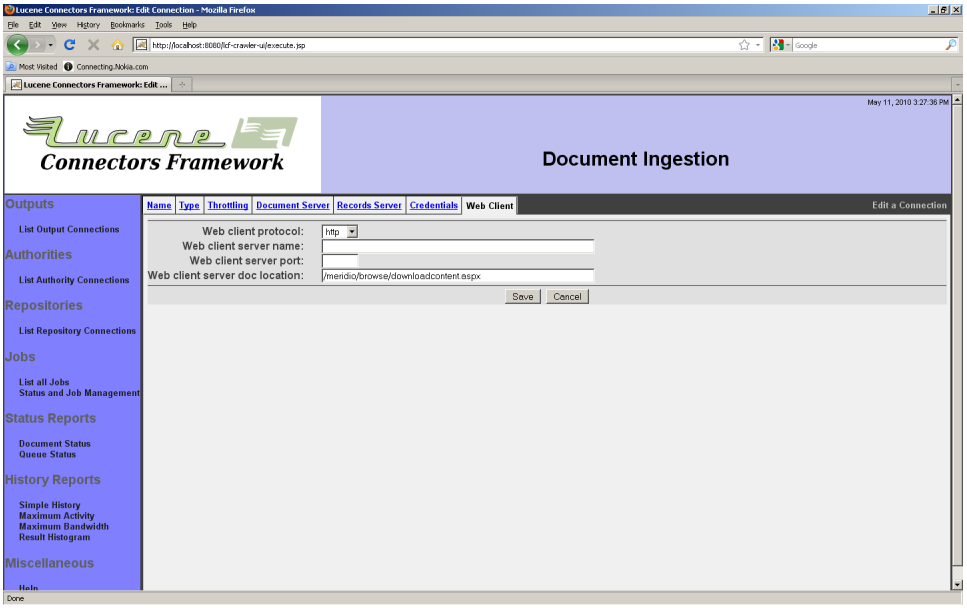
「レコードサーバ」タブを選択すると、次のようなページが表示されます：



プロトコル、サーバ名、ポート番号、Meridioれコートサーバサービスのアドレスを入力してください。プロキシを利用されている場合は、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。

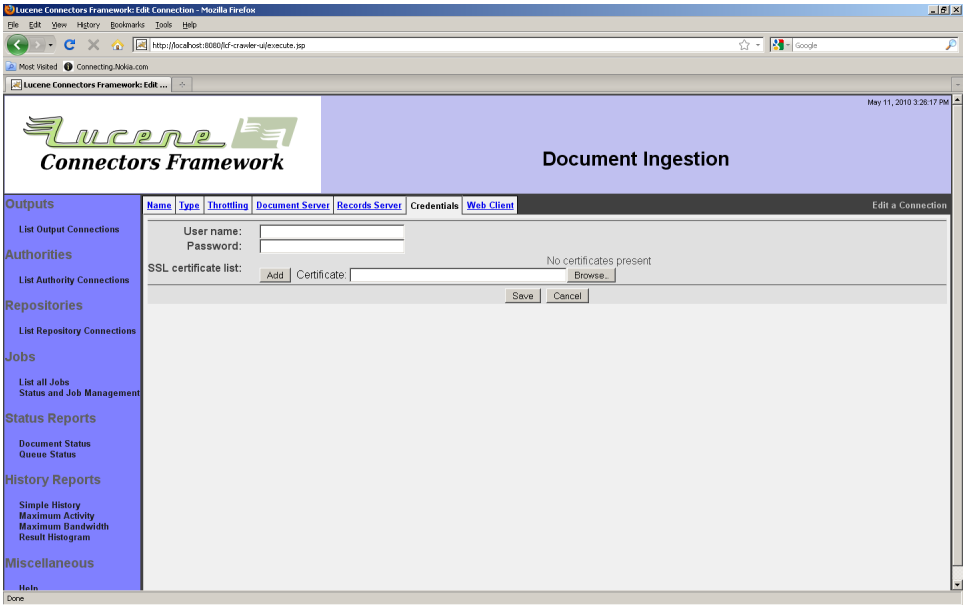
Meridioシステムの場合は異なるサービス毎にサーバを設けることができますが、一般には複数のサービスが同じサーバで動作しています。ただし、コネクションタイプ設定からは異なるサーバを指定することもできます。

「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します：



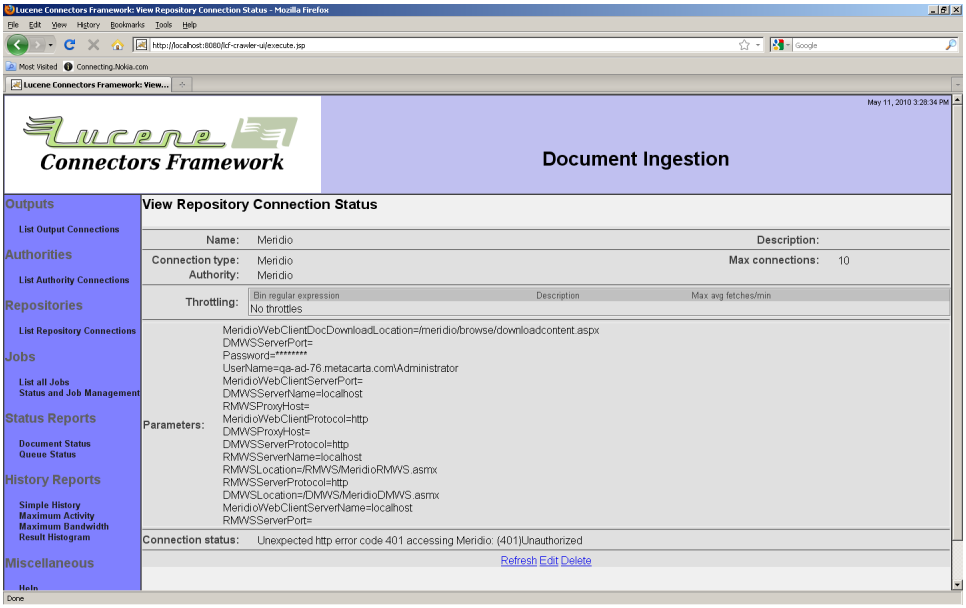
Meridioコネクションwebクライアントタブからは索引を作成したコンテンツ毎にURIを作成します。プロトコルを選択して、サーバ名、ポート番号、Miridio webクライアントサービスのアドレスを入力してください。このサービスからコンテンツを取得しないため、プロキシ情報は不要です。

「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します：



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エアの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します：「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」。

4.12 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。

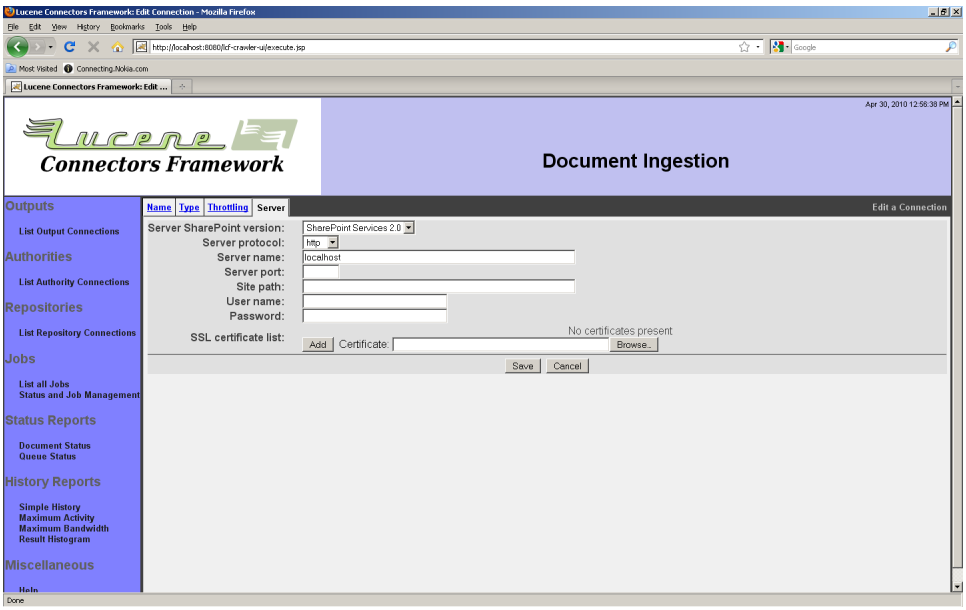
SharePointコネクションタイプは1つのSharePointリポジトリコネクションからルートサイトの明示的なサブサイトを介してすべてのSharePointサイトをアクセスできるように設計されています。大きなSharePointサイトの場合はこのような1つのコネクションからすべて

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アクティブディレクトリで管理されています。アクティブディレクトリ権限を作成していない場合は、「アクティブディレクトリ権限コネクション」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します：



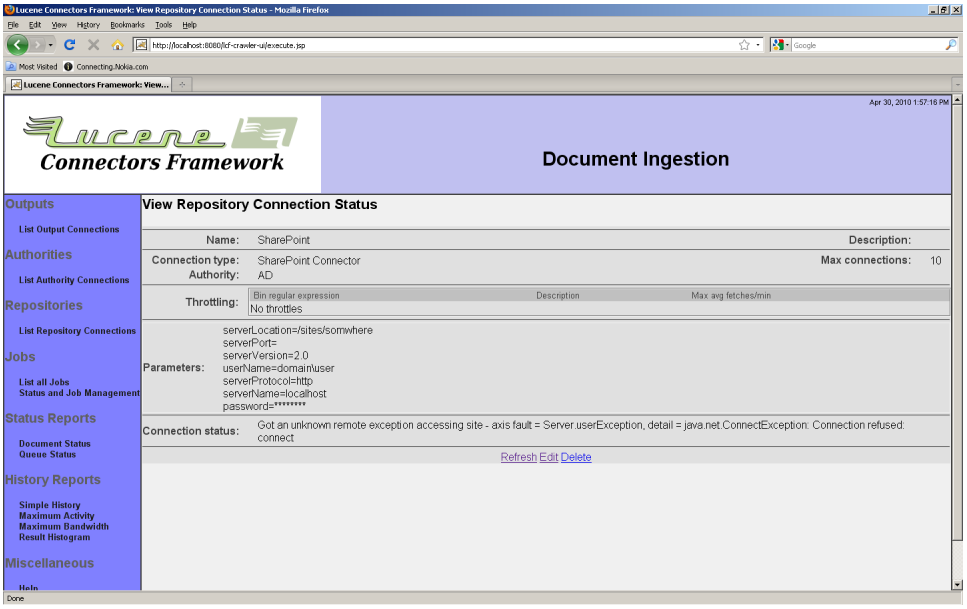
プロダウンリストからSharePointサーバのバージョンを選択してください。間違ったバージョンを選択すると、索引ができなかったり権限情報が正しく取得されない場合があります。

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.asp」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又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します：



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い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エアを修正する必要があります。

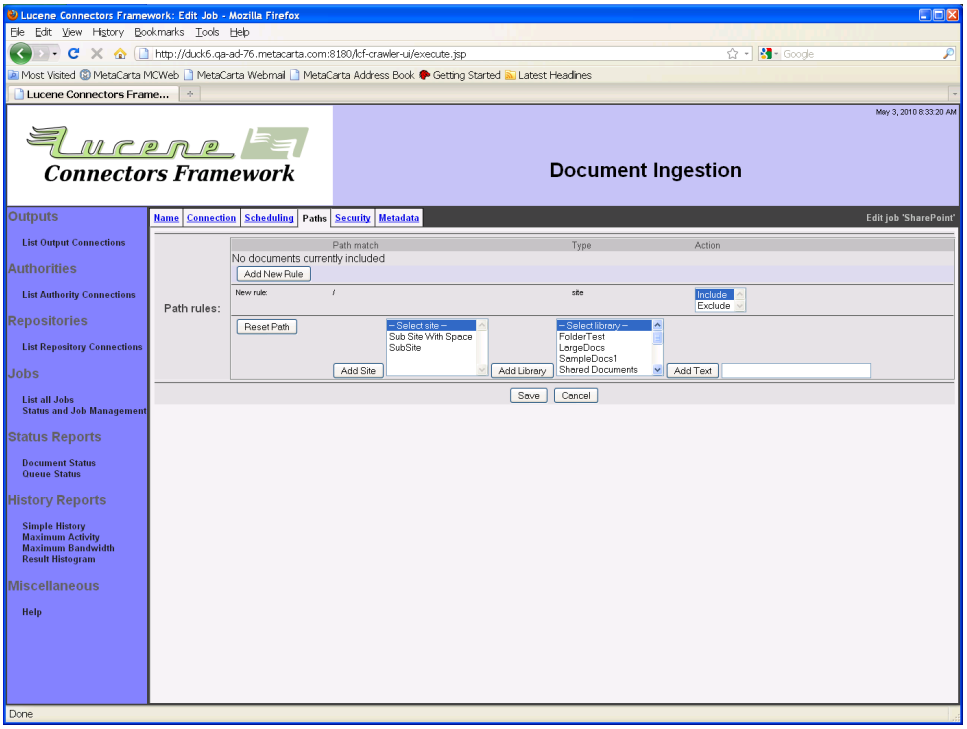
SharePointリポジトリコネクションを選択した場合は、次のようなタブが表示します：「パス」、「セキュリティ」、「メタデータ」。

「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」ルールのみに定義されます。除外ルールには暗黙ルールはありません。

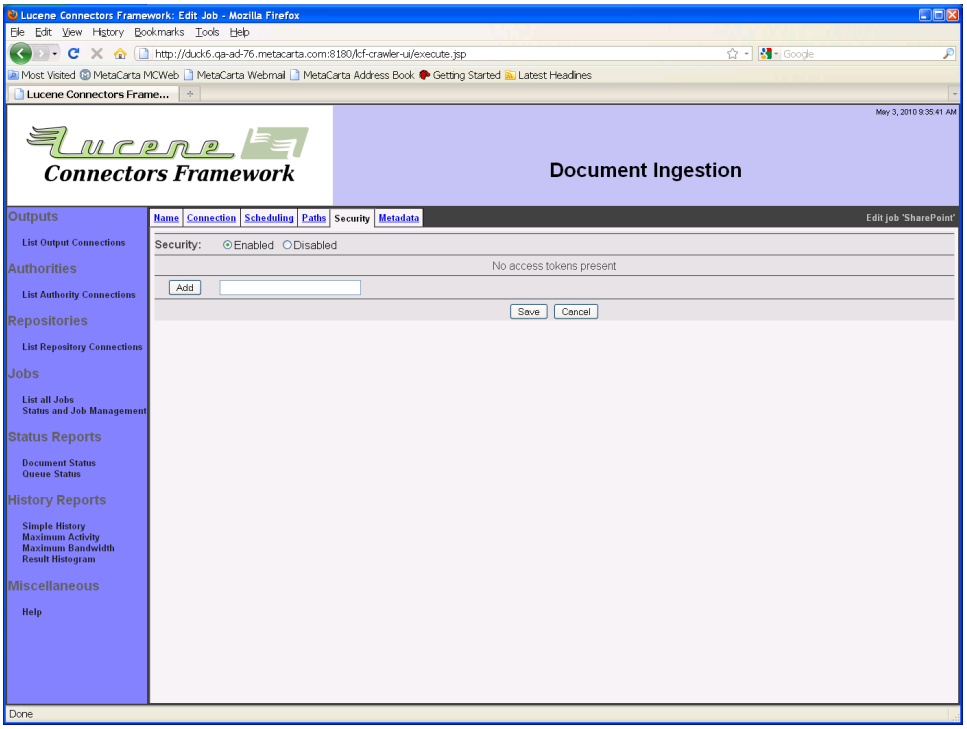
「パス」タブからルールを作成して一覧に追加／挿入することができます。「パス」タブを選択すると次のようなページが表示されます。コネクションが正常に動作していない場合は、プルダウンに表示する項目が少ない場合もあります。



ルールを定義するには、先ず一致するパスを指定します。パスを選択又は入力して「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

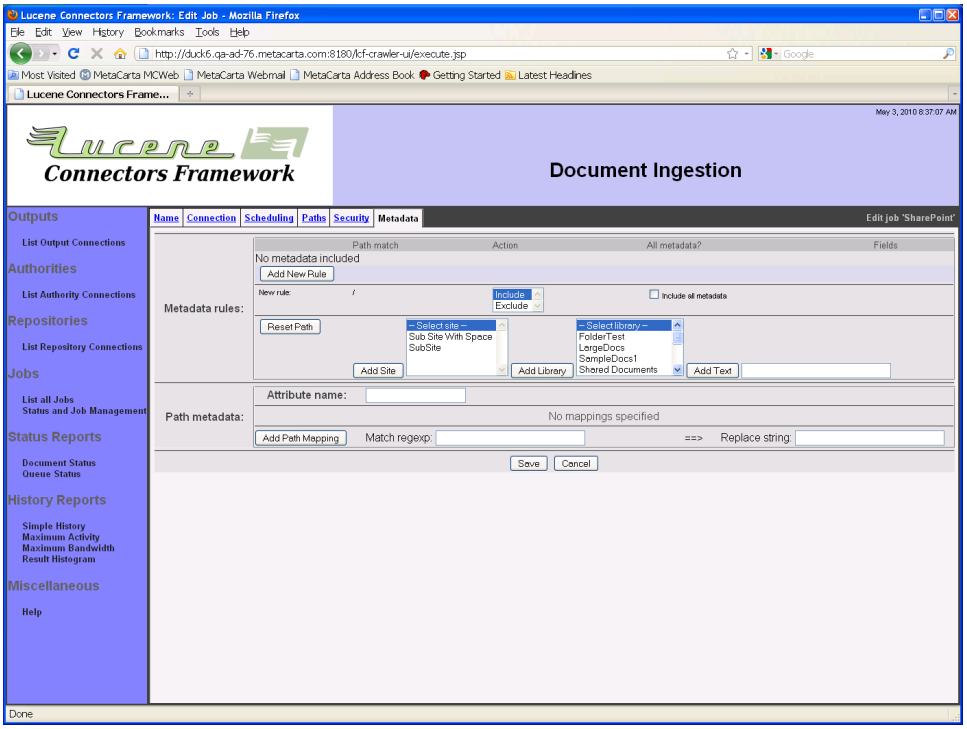
い。次に「含む」又は「除外」用のルールかを選択してください。最後に「新しいルールを追加」ボタンを押下してルールをルール一覧に追加してください。

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します：



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しあ。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します：



「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に含む／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、先ず一致するパスを指定します。パスを選択又は入力して「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」又は「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。変換元と値は格好（「(」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。

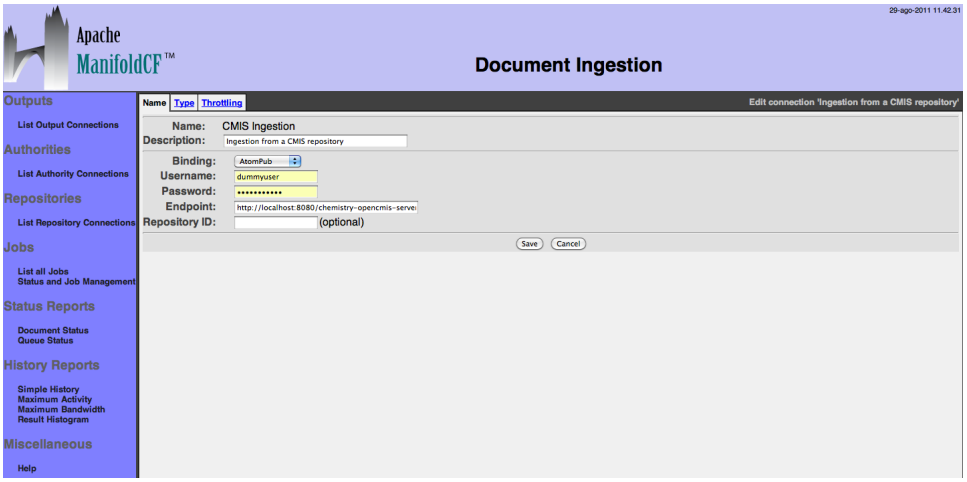
1つ以上のルールが存在する場合は、上から実行され、上のルールの結果は下のルールで変更されます。

4.13 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バイディングプロトコル（AtomPub又は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：

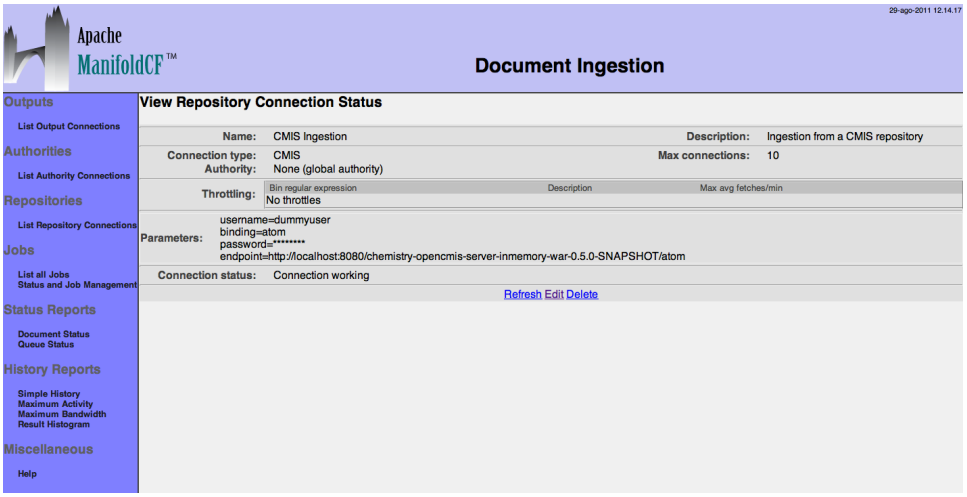
OpenCMISが提供する実際のInMemoryサーバが公開するAtomPubバイディングのエンドポイントは次の通りです：

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

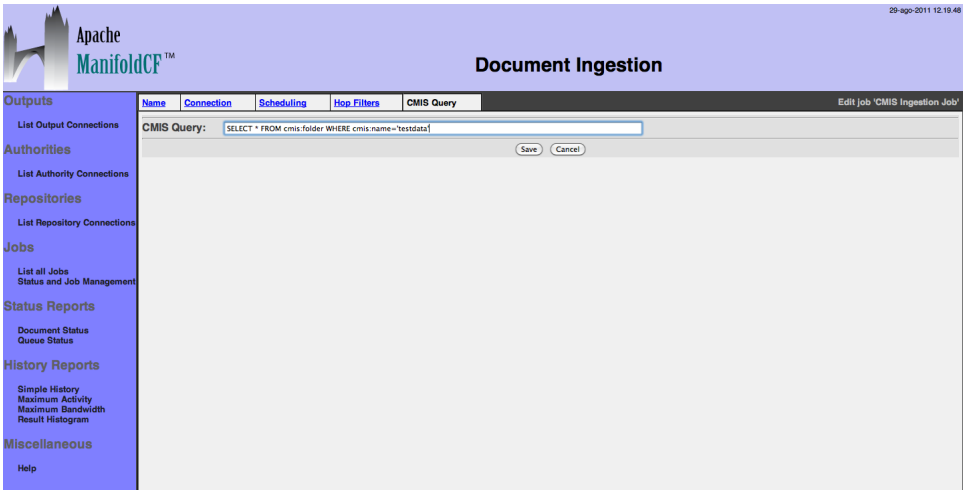
Web Serviceは別のエンドポイントで公開されます：

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：



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」タブが表示します。「CMISクエリー」タブを選択すると次のようなページが表示します：



「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、コンテンツを読込んで処理します。

入力した後には「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：