Copyright © 2013 NTT DATA INTRAMART CORPORATION

目次

- 改訂情報
- はじめに
  - 本書の目的
  - 対象読者
  - 本書の構成
- コンテンツ作成仕様
  - コンテンツ作成対象
  - コンテンツの情報
  - ファセット
  - 閲覧可能権限
- クローラの種類
  - 共通仕様
  - IMBoxクローラ
  - IMBox全件対象クローラ
- テンプレート仕様
  - 表示項目

改訂情報

変更年月日	変更内容
2013-07-01	初版
2013-10-01	第2版 下記を追加・変更しました ■ 「*テンプレート仕様*」のタイプ別表示項目のBroadcastMessageにタグを追加
2014-08-01	第3版 下記を追加・変更しました ■ 「*差分クローリング*」の説明に編集についての記載を追加
2017-08-01	第4版 下記を追加・変更しました ■ 「*クローラの種類*」にupdateCountに関する注意事項を追記

はじめに

項目

- 本書の目的
- 対象読者
- 本書の構成

本書の目的

本書ではIMBoxのクローラの機能概要と仕組の詳細について説明します。

対象読者

次の利用者を対象としています。

- intra-mart Accel PlatformのIMBoxの全文検索機能（IM-ContentsSearch）を管理する運用担当者
- メッセージ種別の追加などのIMBox機能を拡張したコンテンツを全文検索で利用したい開発者

次の内容を理解していることが必須となります。

- intra-mart Accel Platformの仕様
- IM-ContentsSearchの仕様
- IMBoxの仕様

本書の構成

本書は次の構成となっています。

- *コンテンツ作成仕様*
  コンテンツ作成仕様について説明します。
- *クローラの種類*
  クローラの種類について説明します。
- *テンプレート仕様*
  検索結果画面のテンプレートについて説明します。

コラム

用語解説

Storage として使用するディレクトリを %PUBLIC_STORAGE_PATH% と略します。

コンテンツ作成仕様

項目

- コンテンツ作成対象
- コンテンツの情報
  - 標準フィールド
  - 動的フィールド
- ファセット
- 閲覧可能権限

コンテンツ作成対象

IMBoxクローラは各Boxに投稿されたメッセージをスレッドごとにインデックスを作成します。
以下のBoxに投稿されたスレッドが作成対象になります。

- CompanyBox（ブロードキャストメッセージ含む）
- GroupBox
- DirectMessageBox

コンテンツの情報

IMBoxクローラでは、スレッド情報を以下の内容で登録します。

標準フィールド

No	フィールド名	設定値	備考
1	ID	imbox_%スレッドID%	コンテンツを一意に識別するIDです。
2	TYPE	imbox imbox$companybox imbox$companybox$%ボックスコード% imbox$directmessagebox imbox$groupbox imbox$groupbox$%ボックスコード% imbox$broadcast	コンテンツのタイプです。
3	URL	imbox/unitbox/%ボックスコード%/%スレッドID%	スレッド表示画面のURLです。
4	ID_ORIGINAL	スレッドID	コンテンツの元データにおける一意なキーとなるIDです。
5	TITLE	メッセージ内容（200文字まで）	コンテンツのタイトルです。
6	TEXT	メッセージ内容 URI URIタイトル URI説明文 タグ名	コンテンツの主となるテキストデータです。 アンケートは質問内容と返信メッセージのみがフィールド格納対象になります。
7	ATTACHMENT	添付ファイル名 添付ファイル内のテキストデータ	

No	フィールド名	設定値	備考
8	RECORD_DATE	スレッドに投稿された最新のメッセージの投稿日時	

動的フィールド

No	フィールド名	設定値	備考
1	BOX_CD_STRING	ボックスコード	
2	CREATE_THREAD_DATE	スレッド作成日時	
3	CREATE_THREAD_USER_STRING	スレッド作成ユーザ名	
4	DESTINATION_USER_STRING_MLT	DirectMessage宛先ユーザ	
5	MESSAGE_ID_STRING_MLT	メッセージID	
6	MESSAGE_TYPE_CD_STRING	メッセージ種別コード	
7	POST_TYPE_CD_STRING	投稿先種別コード	
8	POST_USER_CD_STRING_MLT	投稿ユーザコード	
9	REPLY_USER_STRING_MLT	返信ユーザ名	
10	TAG_NAME_STRING_MLT	タグ名	

コラム
タイトル（TITLE）、テキスト（TEXT）、添付ファイル（ATTACHMENT）に登録した内容が検索時に検索対象となるフィールドです。

ファセット

IMBoxクローラではコンテンツ作成時に以下のファセット分類を適用しています。

IMBox

IMBoxクローラで作成されたすべてのコンテンツに対して”IMBox”というファセットを適用します。

ブロードキャストメッセージ

ブロードキャストメッセージのコンテンツに対して”ブロードキャストメッセージ”というファセットを適用します。

CompanyBox

CompanyBoxに投稿されたメッセージのスレッドに”CompanyBox”というファセットを適用します。

- 会社名

  “CompanyBox”のファセットを適用したコンテンツに対して、更に会社名のファセットを適用します。

DirectMessageBox

DirectMessageのスレッドに”DirectMessageBox”というファセットを適用します。

GroupBox

GroupBoxに投稿されたメッセージのスレッドに”GroupBox”というファセットを適用します。

- グループ名

  “GroupBox”のファセットを適用したコンテンツに対して、更にグループ名のファセットを適用します。

コラム

IMBoxクローラが全文検索画面に提供するファセットナビゲーションの例

全文検索画面の検索結果にIMBoxのコンテンツがある場合、以下の様なファセットナビゲーションが全文検索画面の左部に表示されます。
リンクを選択することで検索結果の絞込み検索を行うことができます。

閲覧可能権限

IMBoxクローラは作成するコンテンツに対して、コンテンツの閲覧を許可する権限を以下のように設定します。

- CompanyBoxに投稿されたメッセージのスレッド

  投稿先の会社に所属するユーザ
- GroupBox（公開）に投稿されたメッセージのスレッド

  公開先会社に設定している会社に所属するユーザ
- GroupBox（非公開）に投稿されたメッセージのスレッド

  グループに参加中のユーザ
- DirectMessageBoxに投稿されたメッセージのスレッド

  宛先に指定されているユーザ
- ブロードキャストメッセージのスレッド

  IMBoxクローラが実行された時点で有効な会社に所属するユーザ

クローラの種類

項目
- 共通仕様
  - 最終クロール日時
- IMBoxクローラ
  - 差分クローリング
  - 削除クローリング
- IMBox全件対象クローラ
  - 全件作成・再作成処理

共通仕様

ジョブスケジューラ機能のジョブとして以下の２つのジョブを提供しています。

- IMBoxクローラ
- IMBox全件対象クローラ

最終クロール日時

IMBoxクローラとIMBox全件対象クローラでは、処理が正常終了した場合にクローラの最終実行日時をファイルに保存します。
差分クローリング実行時にはこのファイルを参照し、前回の実行日時より後にコンテンツの登録・更新対象となったスレッドを抽出します。
日時のフォーマットは “**yyyy-MM-dd HH:mm:ss**” です。

最終クロール日時を保存するファイルは、クローラジョブの初回正常終了時にPublicStorageの以下のディレクトリに保存されます。

- %PUBLIC_STORAGE_PATH%/products/im_contents_search/store/%テナントID%/last_crawling_date/imbox.json

ただし、削除クローリングが実行された場合、最終クロール日時は “**2000-01-01 00:00:00**” に初期化されます。

初回実行時の動作仕様

初期状態の場合、最終クロール日時を保存するファイルは存在していません。
この場合、既定値として”2000年1月1日0時0分0秒”を最終クロール日時として動作する仕様です。
初回実行時の対象日時を明示的に指定したい場合は、上記ファイルを作成して対象日時を上記の日時フォーマットにて設定してください。

IMBoxクローラ

IMBoxのコンテンツの差分クローリングと削除クローリングを行うジョブです。

ジョブカテゴリ	IM-ContentsSearch クローラ
ジョブID	imbox-crawler
ジョブ名	IMBoxクローラ

IMBoxクローラジョブは標準で以下のジョブネットにより登録されています。

- 差分クローリング（ジョブネットID: crawler-delta-jobnet）

  差分クローリングのジョブネットが実行されたとき、IMBoxクローラの差分クローリングが実行されます。
- 削除クローリング（ジョブネットID: crawler-delete-jobnet）

  削除クローリングのジョブネットが実行されたとき、IMBoxクローラの削除処理が実行されます。

差分クローリング

差分クローリング実行時の登録・更新対象

差分クローリング実行時にコンテンツの登録・更新の対象となるスレッドの条件は以下の通りです。
最終クロール日時以降で以下のいずれかの条件を 1 つでも満たすスレッドについて、コンテンツの登録または更新が行われます。

- メッセージが投稿、編集または削除されたスレッド
- タグが追加または削除されたスレッド
- 宛先追加が行われたDirectMessageBoxのスレッド
- 公開先会社が追加された公開グループ内の全スレッド
- ユーザが新たに参加（招待、および、申請の承認を含む）した非公開グループ内の全スレッド
- ユーザが離脱または除名された非公開グループ内の全スレッド

差分クローリング実行時の実行パラメータ

差分クローリング実行時には差分クローリングのジョブネット、および、IMBoxクローラのジョブ情報に設定されている以下の実行パラメータが利用されます。

キー	説明
crawlingType	クローラの動作タイプを設定します。”DELTA”である場合のみ差分クローリングは動作します。
groupName	処理対象とする検索サーバグループIDを指定します。
maxSegments	最適化の度合いを設定します。
updateCount	Solrサーバへのコミット 1 回あたりに登録するコンテンツの件数
withCommit	処理終了時に更新データを確定するかを設定します。
withOptimize	処理終了時にデータの最適化をするかを設定します。

注意
updateCountの数値が大きい場合、Solrサーバへのコミット1回あたりに登録するコンテンツの件数が増加するため、Solrサーバへの負荷が高くなりパフォーマンスに影響があります。

削除クローリング

IMBoxクローラ、IMBox全件対象クローラが作成したコンテンツをすべて削除します。

削除クローリング実行時の実行パラメータ

削除クローリング実行時には削除クローリングのジョブネット、および、IMBoxクローラのジョブ情報に設定されている以下の実行パラメータが利用されます。

キー	説明
crawlingType	クローラの動作タイプを設定します。”DELETE”である場合のみ削除処理は動作します。
groupName	処理対象とする検索サーバグループIDを指定します。
maxSegments	最適化の度合いを設定します。
withCommit	処理終了時に更新データを確定するかを設定します。
withOptimize	処理終了時にデータの最適化をするかを設定します。

IMBox全件対象クローラ

IMBoxでのコンテンツ作成対象となっている全スレッドの索引化を行うジョブです。

ジョブカテゴリ	IMBox
ジョブID	imbox-all-crawler
ジョブ名	IMBox全件対象クローラ

IMBox全件対象クローラは標準で以下のジョブネットにより登録されています。

- 再作成クローリング（ジョブネットID: crawler-reindex-jobnet）
- IMBox全件対象クローラ（ジョブネットID: imbox-jobnet-all-crawler）

全件作成・再作成処理

IMBoxでのコンテンツ作成対象のコンテンツ情報を全件削除してから、全スレッドの索引化を実行します。
再作成クローリングとIMBox全件対象クローラのジョブネットで実行されるIMBox全件対象クローラジョブの処理に違いはありません。

IMBox全件対象クローラ実行時の実行パラメータ

IMBox全件対象クローラ実行時にはジョブネット、および、ジョブ情報に設定されている以下の実行パラメータが利用されます。

キー	説明
groupName	処理対象とする検索サーバグループIDを指定します。
maxSegments	最適化の度合いを設定します。
queryRangeCount	一度のアクセスでCassandraのメッセージカラムファミリーから取得するデータの件数を設定します。
withCommit	処理終了時に更新データを確定するかを設定します。
withOptimize	処理終了時にデータの最適化をするかを設定します。

テンプレート仕様

項目
- 表示項目
  - 共通表示項目
  - タイプ別表示項目

表示項目

IMBoxクローラが作成するコンテンツのタイプごとに全文検索画面に表示されるコンテンツの項目が異なります。

共通表示項目

それぞれのコンテンツのタイプで共通的に表示される項目は以下です。

- タイトル
- スレッドに投稿された最新のメッセージの投稿日時
- 投稿先のパンくずリスト
- スレッドの投稿ユーザ名
- 要約

  要約表示の指定をして検索した場合に表示されます。
  要約は検索条件の文字列がカラーでハイライトされて表示されます。

タイプ別表示項目

CompanyBox

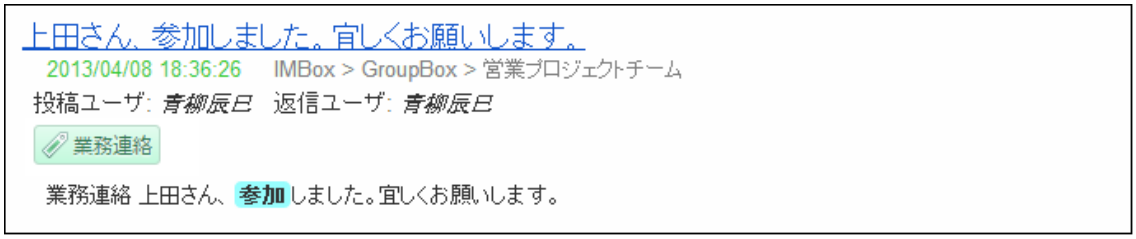
- スレッド内の返信メッセージの投稿ユーザ名
- タグ

検索結果画面例

GroupBox

- スレッド内の返信メッセージの投稿ユーザ名
- タグ（公開グループのみ）

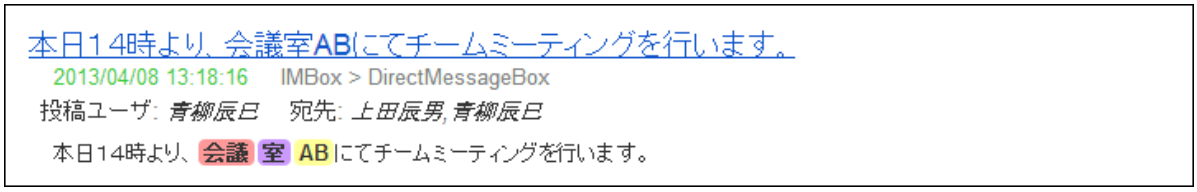
検索結果画面例



DirectMessageBox

- DirectMessageの宛先のユーザ名

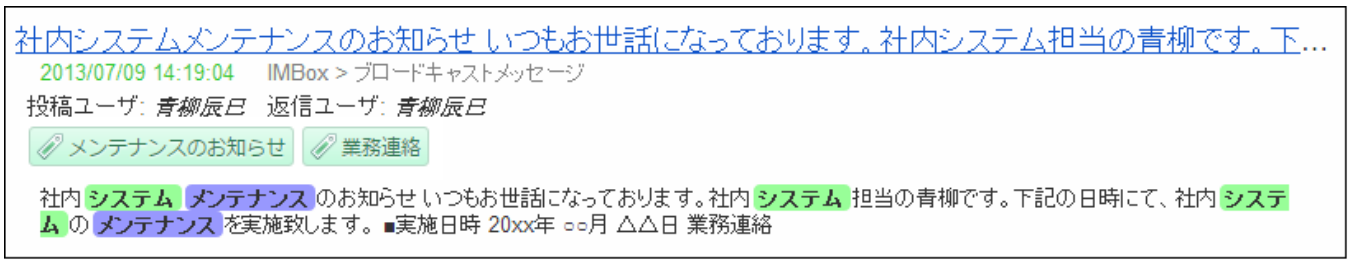
検索結果画面例

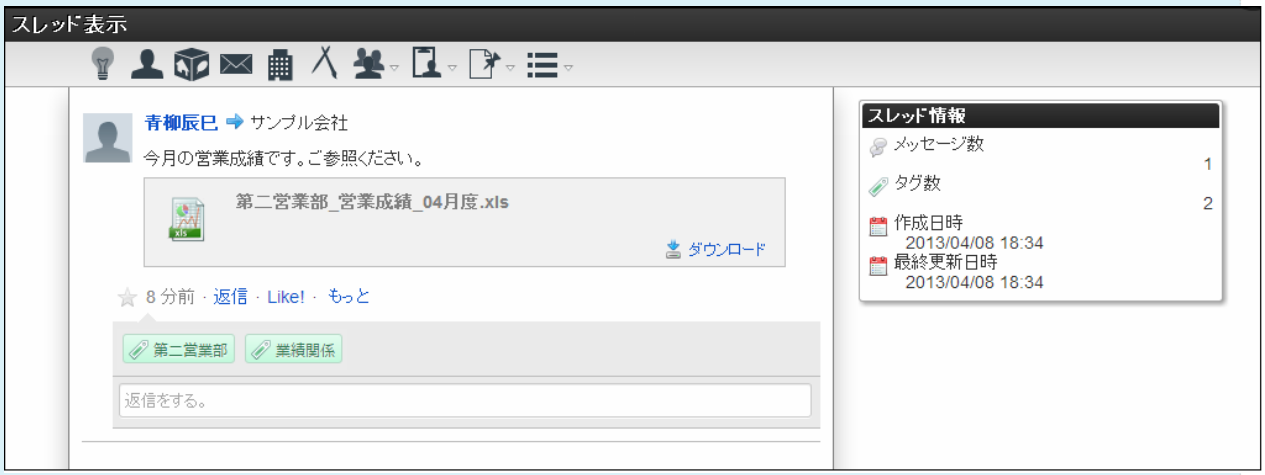


BroadcastMessage

- スレッド内の返信メッセージの投稿ユーザ名
- タグ

検索結果画面例



コラム

検索結果のタイトルリンクをクリックすると、新しいタブにスレッド表示画面が表示されます。