



Copyright © 2018 NTT DATA INTRAMART CORPORATION

目次

- 1. 改訂情報
- 2. はじめに
 - 2.1. 本書の目的
 - 2.2. 対象読者
 - 2.3. 本書の構成
 - 2.4. ライセンスについて
- 3. 共通仕様
 - 3.1. クローラ
 - 3.1.1. 差分クローリング
 - 3.1.2. 削除クローリング
 - 3.1.3. 再作成クローリング
 - 3.1.4. 最終クロール日時
 - 3.2. ファセット
 - 3.3. 閲覧可能権限
 - 3.3.1. 全文検索時の閲覧権限
 - 3.3.2. 各ナレッジコンテンツの閲覧権限
- 4. IM-Wiki全文検索機能
 - 4.1. Wikiページクローラ
 - 4.1.1. 差分クローリング
 - 4.1.2. 削除クローリング
 - 4.1.3. 再作成クローリング
 - 4.2. コンテンツ作成対象
 - 4.3. コンテンツの情報
 - 4.3.1. 標準フィールド
 - 4.3.2. 動的フィールド
 - 4.4. ファセット
 - 4.5. 閲覧可能権限
 - 4.6. 検索結果表示項目
- 5. ワークブック全文検索機能
 - 5.1. ワークブッククローラ
 - 5.1.1. 差分クローリング
 - 5.1.2. 削除クローリング
 - 5.1.3. 再作成クローリング
 - 5.2. コンテンツ作成対象
 - 5.3. コンテンツの情報
 - 5.3.1. 標準フィールド
 - 5.3.2. 動的フィールド
 - 5.4. ファセット
 - 5.5. 閲覧可能権限
 - 5.6. 検索結果表示項目

変更年月日　変更内容

2018-12-01 初版

項目

- 本書の目的
- 対象読者
- 本書の構成
- ライセンスについて

本書の目的

本書ではIM-Knowledgeの各ナレッジコンテンツの全文検索の機能概要と仕組み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。

- IM-Knowledgeの全文検索機能（IM-ContentsSearch）を管理する運用担当者

次の内容を理解していることが必須です。

- intra-mart Accel Platformの仕様
- IM-ContentsSearchの仕様
- IM-Knowledgeの仕様
- IM-Wikiの仕様（IM-Wiki機能の全文検索を利用する場合）
- IM-Spreadsheetの仕様（IM-Spreadsheet機能の全文検索を利用する場合）

本書の構成

本書は次の構成です。

- **共通仕様**

IM-Knowledgeの全文検索機能の共通的な仕様について説明します。

- **IM-Wiki全文検索機能**

IM-WikiのWikiページの全文検索の仕様について説明します。

- **ワークブック全文検索機能**

IM-Spreadsheetのワークブックの全文検索の仕様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。

IM-ContentsSearchの用語である「コンテンツ」（全文検索の対象となるデータを表す単位）は単に「コンテンツ」と表記します。

IM-Knowledgeの用語である「コンテンツ」（ナレッジとして保存する情報）はIM-ContentsSearchのコンテンツと区別するため「ナレッジコンテンツ」と表記します。

IM-ContentsSearchのライセンスについての詳細は以下のドキュメントを参照してください。

- 「[IM-ContentsSearch for Accel Platform リリースノート](#)」 - 「[ライセンスについて](#)」

IM-Spreadsheetのライセンスについての詳細は以下のドキュメントを参照してください。

- 「[IM-Spreadsheet 利用ガイド](#)」 - 「[ライセンスについて](#)」

この章では、IM-Knowledgeの各ナレッジコンテンツの全文検索機能における共通的な仕様について説明します。

項目

- クローラ
 - 差分クローリング
 - 削除クローリング
 - 再作成クローリング
 - 最終クロール日時
- ファセット
- 閲覧可能権限
 - 全文検索時の閲覧権限
 - 各ナレッジコンテンツの閲覧権限

クローラ

IM-Knowledgeのクローラはジョブスケジューラ機能のジョブとしてナレッジコンテンツの種別ごとにジョブを提供しています。

それぞれのジョブはテナント環境セットアップ時に「IM-ContentsSearchクローラ」ジョブカテゴリ（カテゴリID：crawler-job-category）に登録されます。

ナレッジコンテンツ

種別	ジョブ名	ジョブID
IM-Wiki	Wikiページクローラ	wiki-crawler
IM-Spreadsheet	ワークブッククローラ	spreadsheet-crawler

上記のジョブは標準で以下のジョブネットに追加されます。

ジョブネット名	ジョブネットID
差分クローリング	crawler-delta-jobnet
削除クローリング	crawler-delete-jobnet
再作成クローリング	crawler-reindex-jobnet

差分クローリング

差分クローリングのジョブネットが実行されたとき、各ナレッジコンテンツのクローラの差分クローリングが実行されます。

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、および各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が動作します。

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。
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

削除クローリングのジョブネットが実行されたとき、各ナレッジコンテンツのクローラのインデックス削除処理が実行されます。

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、および各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。

再作成クローリング

再作成クローリングのジョブネットが実行されたとき、各ナレッジコンテンツのクローラの再作成処理が実行されます。

再作成クローリング実行時の実行パラメータ

再作成クローリング実行時には再作成クローリングのジョブネット、および各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。"REINDEX"である場合のみ再作成クローリングが動作します。
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。

キー	説明
----	----

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。

最終クロール日時

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に保存されます。

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“**1970-01-01 00:00:00**”に初期化されて保存されます。

ファイルの保存先はクローラごとに異なります。

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“1970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。

ファセット

各ナレッジコンテンツのクローラでは、作成されたすべてのコンテンツに対して「**Knowledge**」というファセットを適用します。

その他のファセットは各ナレッジコンテンツの種類ごとに適用されます。各ナレッジコンテンツのクローラが適用するファセットについては、各機能の「ファセット」の項を参照してください。

閲覧可能権限

全文検索時の閲覧権限

各ナレッジコンテンツのクローラはジョブ実行日時の時点での権限を判定し、コンテンツに対して閲覧権限を適用します。

全文検索で閲覧可能なコンテンツはナレッジグループの認可設定の権限で判定されます。

Solrサーバに保存されるインデックスにはナレッジグループに設定されている認可の詳細な設定は保持しておらず、全文検索に表示されるかどうかの判定はインデックスに保存されているナレッジグループと、検索したユーザが閲覧可能なナレッジグループを比較して全文検索の閲覧権限を判定しています。

従って、ナレッジコンテンツのナレッジグループを別のナレッジグループに変更した場合は再度クローリングする必要がありますが、ナレッジグループに設定されている認可設定を変更した場合のみでは再度クローリングを実行する必要はありません。

各ナレッジコンテンツの閲覧権限

検索結果に表示されたコンテンツの画面表示・処理を行うことができるかどうかはIM-Knowledgeのナレッジグループの認可設定の権限で判定されます。

IM-Knowledgeに作成されたWikiページの全文検索機能に関する仕様について説明します。

項目

- Wikiページクローラ
 - 差分クローリング
 - 削除クローリング
 - 再作成クローリング
- 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
- 検索結果表示項目

Wikiページクローラ

Wikiページのクローリング処理を行うジョブです。

最終クロール日時保存場所

Wikiページクローラは最終クロール日時をPublicStorageの以下のファイルに保存します。

```
<%PUBLIC_STORAGE_PATH%/products/im_contents_search/store/%テナント  
ID%/last_crawling_date/imkb$wiki.json>
```

差分クローリング

差分クローリング時の登録・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Wikiページ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つ以上満たすWikiページについて、コンテンツの登録または更新処理が実行されます。

- 条件1：新たに登録されたWikiページ
- 条件2：更新されたWikiページ
- 条件3：削除されたWikiページ
- 条件4：タグが付与されたWiki内の全ページ
- 条件5：ナレッジグループが更新されたWiki内の全ページ

削除クローリング

Wikiページクローラが作成したコンテンツをすべて削除します。

また、最終クロール日時を“**1970-01-01 00:00:00**”に初期化して保存します。

再作成クローリング

コンテンツ作成対象

WikiページクローラはすべてのWikiページがコンテンツ作成対象です。 Wikiページ1つごとにインデックスを作成します。

コンテンツの情報

WikiページクローラではWikiページ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	imkb_%WikiページID%	コンテンツを一意に識別するIDです。
2	TYPE	imkb imkb\$wiki imkb\$wiki\$%ナレッジグループID%	コンテンツのタイプです。
3	URL	knowledge/contents/wiki/%コンテン ツID%/%Wikiページタイトル%	Wikiページ表示画面のURLです。
4	ID_ORIGINAL	Wikiページ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	Wikiページタイトル	コンテンツのタイトルです。
6	TEXT	Wikiページ本文	コンテンツの主となるテキストデータです。
7	ATTACHMENT	添付ファイル名 添付ファイル内のテキストデータ	
8	RECORD_DATE	Wikiページ最終更新日時	



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。

動的フィールド

No	フィールド名	設定値	備考
1	GROUP_DESCRIPTION_STRING	ナレッジグループ説明	JSON形式、多言語情報
2	GROUP_NAME_STRING	ナレッジグループ名	JSON形式、多言語情報
3	CREATE_USER_CD_STRING	Wikiページ作成ユーザコード	
4	CREATE_USER_NAME_STRING	Wikiページ作成ユーザ名	
5	CREATE_DATE_DATE	Wikiページ作成日時	
6	RECORD_USER_CD_STRING	Wikiページ最終更新ユーザコード	

No	フィールド名	設定値	備考
7	RECORD_USER_NAME_STRING	Wikiページ最終更新ユーザ名	
8	CONTENTS_COMMENT_STRING	Wikiページ更新コメント	
9	LATEST_VERSION_INTEGER	Wikiページの最終更新バージョン番号	
10	FORMAT_TYPE_STRING	Wikiページのテキスト形式	
11	TAG_STRING_MLT	タグ情報	JSON形式

ファセット

Wikiページ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。

Wiki

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

ナレッジグループ名

「Wiki」のファセットを適用したコンテンツに対して、更にナレッジグループ名のファセットを適用します。

i コラム

IM-Wiki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Wikiページクローラ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。

絞り込み条件

- ▼ Knowledge (22)
- ▼ Wiki (22)
- サンプルデータ (22)

閲覧可能権限

作成されたコンテンツの閲覧を許可する権限は、ナレッジグループに設定されているアクセス権と同じ権限が適用されます。

ユーザが検索を実行したときに検索対象となるWikiページは、そのユーザに許可されているナレッジグループのアクセス権で「管理」・「編集」・「参照」のいずれかが許可されているナレッジグループのWikiページです。

検索結果表示項目

Wikiページクローラが作成したコンテンツの検索結果で全文検索画面に表示されるコンテンツの表示項目は以下です。

- タイトル (Wikiページタイトル)
- Wikiページ最終更新日時
- ナレッジグループ名

- Wikiページ作成者名
- Wikiページ更新者名
- コンテンツに付与されているタグ名
- 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。

検索結果画面例

議事録サンプル - 定例会議議事録20180801

2018/10/02 19:11:24 Knowledge > Wiki > サンプルデータ

グループ名 : サンプルデータ 説明 :

作成者 : 青柳辰巳 更新者 : 青柳辰巳

data"]} 定例会議 [議事] 錄 20180801 # サンプル課会議 [議事] 錄 ## 日時・場所・参加者 * 日時 * 8月1日(水) 09:30~10:00 * 場所 * 第1会議室 * 参加者(敬称略)

i コラム

検索結果のタイトルリンクをクリックすると、新しいウィンドウで「Wikiページ表示」画面を表示します。

IM-Knowledgeに作成されたIM-Spreadsheetのスプレッドシート（ワークブック）の全文検索機能に関する仕様について説明します。

項目

- ワークブッククローラ
 - 差分クローリング
 - 削除クローリング
 - 再作成クローリング
- 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
- 検索結果表示項目

ワークブッククローラ

ワークブックのクローリング処理を行うジョブです。

最終クロール日時保存場所

ワークブッククローラは最終クロール日時をPublicStorageの以下のファイルに保存します。

```
<%PUBLIC_STORAGE_PATH%/products/im_contents_search/store/%テナント  
ID%/last_crawling_date/imkb$spread.json>
```

差分クローリング

差分クローリング時の登録・更新対象

差分クローリング実行時にコンテンツの登録・更新の対象となるワークブック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つ以上満たすワークブックについて、コンテンツの登録または更新処理が実行されます。

- 条件1：新たに登録されたワークブック
- 条件2：更新されたワークブック
- 条件3：削除されたワークブック
- 条件4：タグが付与されたワークブック
- 条件5：タグが更新されたワークブック
- 条件6：ナレッジグループが更新されたワークブック

削除クローリング

ワークブッククローラが作成したコンテンツをすべて削除します。

また、最終クロール日時を“**1970-01-01 00:00:00**”に初期化して保存します。

ワークブッククローラが作成したコンテンツをすべて削除し、作成対象のワークブックのコンテンツを再度登録します。

コンテンツ作成対象

ワークブッククローラはすべてのワークブックがコンテンツ作成対象です。

ワークブック1つごとにインデックスを作成します。



注意

ワークブッククローラが作成するコンテンツはIM-Knowledgeのナレッジコンテンツとして保存されているワークブックが対象です。

SpreadJS(Spread.Sheets、Spread.Views)を利用して作成したアプリケーション上のデータは、ワークブッククローラが作成するコンテンツに含まれません。

コンテンツの情報

ワークブッククローラではワークブック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	imkb_%ワークブックID%	コンテンツを一意に識別するIDです。
2	TYPE	imkb imkb\$spread imkb\$spread\$%ナレッジグループ ID%	コンテンツのタイプです。
3	URL	knowledge/contents/spreadsheet/% ワークブック表示画面のURLです。 ワークブックID%	ワークブック表示画面のURLです。
4	ID_ORIGINAL	ワークブック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	ワークブックタイトル	コンテンツのタイトルです。
6	TEXT	シート名 セル内の文字列	コンテンツの主となるテキストデータです。
7	ATTACHMENT	-	
8	RECORD_DATE	ワークブック最終更新日時	



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。



注意

検索対象となるセルの文字列について

セル内の文字列はコンテンツ作成対象（全文検索対象）ですが、書式設定を変更しているセルは変換前の文字列が検索対象となる場合があります。

動的フィールド

No	フィールド名	設定値	備考
1	GROUP_DESCRIPTION_STRING	ナレッジグループ説明	JSON形式、多言語情報
2	GROUP_NAME_STRING	ナレッジグループ名	JSON形式、多言語情報
3	CREATE_USER_CD_STRING	ワークブック作成ユーザコード	
4	CREATE_USER_NAME_STRING	ワークブック作成ユーザ名	
5	CREATE_DATE_DATE	ワークブック作成日時	
6	RECORD_USER_CD_STRING	ワークブック最終更新ユーザコード	
7	RECORD_USER_NAME_STRING	ワークブック最終更新ユーザ名	
8	CONTENTS_DESCRIPTION_STRING	ワークブック説明	
9	CONTENTS_COMMENT_STRING	ワークブック更新コメント	
10	LATEST_VERSION_INTEGER	ワークブックの最終更新バージョン番号	
11	FORMAT_TYPE_STRING	ワークブックのテキスト形式	
12	TAG_STRING_MLT	タグ情報	JSON形式

ファセット

ワークブック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。

ワークブック

ワークブッククローラで作成されたすべてのコンテンツに対して「ワークブック」というファセットを適用します。

ナレッジグループ名

「ワークブック」のファセットを適用したコンテンツに対して、更にナレッジグループ名のファセットを適用します。



コラム

ワークブック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ワークブッククローラ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。

Narrow Down Conditions

▼ Knowledge (1)

 ▼ Workbook (1)

 □ Sample data (1)

閲覧可能権限

作成されたコンテンツの閲覧を許可する権限は、ナレッジグループに設定されているアクセス権と同じ権限が適用されます。

ユーザが検索を実行したときに検索対象となるワークブックは、そのユーザに許可されているナレッジグループのアクセス権で「管理」・「編集」・「参照」のいずれかが許可されているナレッジグループのワークブックです。

検索結果表示項目

ワークブッククローラが作成したコンテンツの検索結果で全文検索画面に表示されるコンテンツの表示項目は以下です。

- タイトル（ワークブックタイトル）
- ワークブック最終更新日時
- ナレッジグループ名
- ナレッジグループ説明
- ワークブック作成者名
- ワークブック更新者名
- コンテンツに付与されているタグ名

検索結果画面例

[Employee List](#)

2018/10/04 18:24:48 Knowledge > Workbook > Sample data

Group Name : Sample data Description : This is our employee list

Creation User : Tatsumi Aoyagi Update User : Tatsumi Aoyagi

TAG

MANAGEM...



コラム

検索結果のタイトルリンクをクリックすると、新しいウィンドウで「ワークブック表示」画面を表示します。

The screenshot shows a Microsoft Excel spreadsheet titled "Employee List". The table has 18 rows of data. The columns are labeled A through P. The first row contains the column headers: id, first_name, last_name, email, gender, and ip_address. The data rows show various employee details such as Keene, Grievson, Ferdie, Churchward, Billy, Reddlesden, Heida, Walkey, Trude, Krauss, Ezra, Curston, Blisse, Louden, West, MacKenny, Mattheus, Colcutt, Giordano, Corsor, Kipp, De Isle, Ebba, Holston, Melesa, Dosedale, Weylin, and Nessa, St. Quintin. The "gender" column uses dropdown arrows to select values like Male and Female.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	
1	Employee List															
2																
3	id	first_name	last_name	email	gender	ip_address										
4	1	Keene	Grievson	kgrievson0@furl.net	Male	40.174.40.77										
5	2	Ferdie	Churchward	fchurchward1@answers.com	Male	220.189.163.113										
6	3	Billy	Reddlesden	breddlesden2@cbslocal.com	Male	68.90.44.35										
7	4	Heida	Walkey	hwalkey3@jimdo.com	Female	110.209.204.224										
8	5	Trude	Krauss	takrauss4@addthis.com	Female	167.179.169.118										
9	6	Ezra	Curston	ecurston5@studiodpress.com	Male	181.75.66.175										
10	7	Blisse	Louden	blouden6@mayoclinic.com	Female	208.225.27.251										
11	8	West	MacKenny	wmackenny7@goo.gl	Male	79.77.30.227										
12	9	Mattheus	Colcutt	mcolcutt8@clickbank.net	Male	185.83.87.10										
13	10	Giordano	Corsor	g corsor9@europa.eu	Male	235.227.108.23										
14	11	Kipp	De Isle	kdeislea@elegantthemes.com	Female	146.151.28.4										
15	12	Ebba	Holston	eholstonb@amazon.de	Female	56.142.56.35										
16	13	Melesa	Dosedale	mdosedalec@1und1.de	Female	36.95.152.181										
17	14	Weylin	Godwyn	wgodwynd@seattletimes.com	Male	240.160.122.252										
18	15	Nessa	St. Quintin	nstquintine@hatena.ne.jp	Female	130.227.133.177										