



Coveo Platform 7.0

CKJT Language Improved Relevancy

Notice

The content in this document represents the current view of Coveo as of the date of publication. Because Coveo continually responds to changing market conditions, information in this document is subject to change without notice. For the latest documentation, visit our website at www.coveo.com.

© Coveo Solutions Inc., 2018

Coveo is a trademark of Coveo Solutions Inc. This document is protected by intellectual property laws and is subject to all restrictions specified in the Coveo Customer Agreement.

Document part number: PM-18-26-06-JA

Publication date: 6/26/2018

Table of Contents

1. 中国語日本語韓国語タイ語関連性の向上	1
-----------------------------	---

1. 中国語日本語韓国語タイ語関連性の向上

Coveo プラットフォーム 7 は中国語、日本語、韓国語、タイ語 (CJKT: これらの言語の総称) をサポートしており、現在は改善された関連性をオフターしています。

これらの言語では、言葉を分ける際にスペースを使用しません。以前まで、Coveo Enterprise Search (CES) はそれぞれの文字を単語のように索引付けをし、文字のペアを使って検索を行っていました。この方法によってエンドユーザーによる検索が可能となっていました、検索結果やランキングには最適なものではありませんでした。

新しい索引付けを用いた方法では、Coveo Enterprise Search (CES) は証明済みの言語トークナイザを使用し、CJKT 単語と呼ばれる分離不可能な文字のグループ内の式を識別、分離します。それぞれのCJKT単語は、通常の単語として索引付けされます。このようにCJKT単語の意味は保持されて、文字の一つ一つではなく単語に基づいてランキングが作られるので関連性が向上します。

例: 検索ボックスには中国語、日本語、韓国語、タイ語のキーワードの入力が可能で、検索結果の中から関連した文書を取得し、検索結果のタイトルとCJKTキーワードのハイライトを見たり、次の中国語の例のように抜粋します。



注意:

- **CES 7.0.6547+ (2014年3月)** 中国語、日本語、韓国語の関連性を改善。
- **CES 7.0.6424+ (2014年2月)** タイ語の関連性を改善。
- **CES 7.0.6547+ (2014年3月)** 新しく作成された索引では、新規の関連性の高いCJKT索引付け方法を使用します。CES 7.0.6547より前のバージョンからCES 7.0.6547+のバージョンにCESをアップグレードすると、既存のインデックスは、デフォルトで元のCJKT索引方法を継続して使用します。新しいCJKT索引方法の使用を希望する場合は [Coveo サポート](#) にお問い合わせください。

以下のような例では、類似の大文字のグループ(例.: TTT) CJKT単語を表示し、同時に異なる小文字のグループがCJKT単語でない単語を表示します。

例: 簡体中国語の表現などCoveoは多くの言語をサポートしており以下のように分類されます:

オリジナルの表現: coveo支持多国语言
 代表: abc TTTUUU
 場所: abc 支持coveo
 TTT 支持 (サポート)を表す
 UUU 多国语言 (多言語)を表す

サポートされている機能	説明
索引	多言語固有のUnicode文字セットとエンコーディングに基づいたCJKTコンテンツの自動検出 他の言語と同様に、CJKT単語はどの文書が表示されるか、各文書に表示される場所を識別するために索引付けされます。
検索	<p>照会時に、CJKT式はCJKT単語に分割され、検索結果は全てのCJKT単語を含む全ての文書を提示します。検索インターフェースでは、検索されたCJKT単語が検索結果のタイトルと抜粋で強調表示されます。エンドユーザーは、CJKT以外の単語または用語を混在させたCJKT単語を検索することができます。</p> <p>例:</p> <p>入力された質問: abcTTTUUUdef 変換された質問: (abc TTT UUU def)</p> <p>注意:返却された文書は、その他の言語と同じプロセス及び基準を使用してランク付けされます。</p>

サポートされている機能	説明
接頭辞と演算子	<p>Coveo検索インターフェースの検索ボックスでは、エンドユーザーは検索接頭辞とCJKT式の演算子を使用できます。ブーリアン演算子は英語での表記が必須です(AND, OR, NEAR, NOT).</p> <p>例: エンドユーザーは、単語とCJKT式の間でOR演算子を使用できます:</p> <p><u>型付きクエリ:</u> abc OR TTTUUU</p> <p><u>変換されたクエリ:</u> (abc OR (TTT UUU))</p> <p>例: エンドユーザーは、単語とCJKT式の間でNEAR演算子を使用できます:</p> <p><u>型付きクエリ:</u> abc NEAR TTTUUU</p> <p><u>変換されたクエリ:</u> (abc NEAR "TTT UUU")</p> <p><u>型付きクエリ:</u> r-cTTTUUU NEAR def</p> <p><u>変換されたクエリ:</u> ("r c TTT UUU" NEAR def)</p> <p>注意: NEAR演算子は、単語やフレーズのマッチングをサポートしますが、サブ表現はマッチングできません。</p> <p>例: エンドユーザーは、CJKT式の前にある完全なフレーズ一致に展開するNOT演算子またはマイナス演算子を使用できます:</p> <p><u>型付きクエリ:</u> NOT TTTUUU</p> <p><u>変換されたクエリ:</u> NOT "TTT UUU"</p> <p><u>型付きクエリ:</u> -TTTUUU</p> <p><u>変換されたクエリ:</u> -"TTT UUU"</p> <p>例: CJKTにすてミングは適用されませんが、エンドユーザーは引き続き完全一致を使用するか、(+) 数値(#) 演算子をCJKT式の前に置いて式を完全な句として展開します。演算子は取り除かれます。</p> <p><u>型付きクエリ:</u> +TTTUUU</p> <p><u>変換されたクエリ:</u> "TTT UUU"</p> <p><u>型付きクエリ:</u> #TTTUUU</p> <p><u>変換されたクエリ:</u> "TTT UUU"</p>

サポートされている機能	説明
フレーズ検索	<p>Coveo検索インターフェースの検索ボックスで、エンドユーザーは特定のCJKTフレーズを検索できます。フレーズセマンティックは保存されます。</p> <p>例: エンドユーザーは二重引用符を使用して、式を完全に一致させることができます:</p> <p><u>型付きクエリ:</u> abc"TTTUUUdef"</p> <p><u>変換されたクエリ:</u> abc "TTT UUU def"</p> <p>単語以外の文字は、周囲の文字の正確なフレーズを生成します。</p> <p>例: ダッシュ(-)が存在すると、完全なフレーズ一致に変換されます:</p> <p><u>型付きクエリ:</u> TTTUUU-VVV</p> <p><u>変換されたクエリ:</u> "TTT UUU VVV"</p>
シソーラス	<p>A Coveoの管理者rはシソーラスのエントリにCJKT式を入力してクエリを展開できます。</p> <p>注意: CJKTシソーラス項目はCJKT単語に適用されるので、CJKT式とCJKT単語は同等と見なされます。</p> <p>例: シソーラスのエントリにTTTUUUまたはTTT UUUを入力することは同効果をもたらします。</p>
フィールドクエリ	<p>タイ語の式はフィールドクエリで使用でき、正確なフレーズ一致に変換されるので、フィールド内の一致はより正確です。</p> <p>例:</p> <p><u>型付きクエリ:</u> @field=abcTTTUUU</p> <p><u>変換されたクエリ:</u> @field="abc TTT UUU"</p> <p><u>型付きクエリ:</u> @field=(abc,TTTUUU)</p> <p><u>変換されたクエリ:</u> @field=(abc, "TTT UUU")</p> <p><u>型付きクエリ:</u> @field=(abc,"TTTUUU")</p> <p><u>変換されたクエリ:</u> @field=(abc, "TTT UUU")</p>
ストップワード	<p>A Coveoの管理者はストップワードとしてCJKT単語を含めることができます。</p>
もしかして	<p>Word Corrector Lexicon (WCL) はCJKTの単語をサポートしているので、誤ったCJKT式はクエリ時にもしかしての提案につながります。</p>

注意: 中国語、日本語、韓国語、タイ語では、ワイルドカードの使用はサポートされていません。