



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-KO

Publication date: 6/26/2018

Table of Contents

1. 향상된 중국어 일본어 한국어 태국어의 검색능력	1
------------------------------------	---

1. 향상된 중국어 일본어 한국어 태국어의 검색능력

Coveo 플랫폼 7은 항상 중국어, 일본어, 한국어, 태국어 (CJKT)를 지원해왔지만, 이제 더 나아진 이 언어들의 검색능력을 제공한다.

이 언어들은 각 단어마다 띄어쓰기를 하지 않는다. 앞서서의 Coveo Enterprise Search (CES)는 각각의 글자를 따로따로 하나의 단어인양 분류를 하여, 글자의 조합으로 검색이 이뤄지게 했었다. 이 방식은 사용자들이 원하는 내용을 찾을 수는 있게 하였지만, 검색결과와 정확성과 순위에는 적합하지 않았다.

새로 도입한 색인 방법에 따라 Coveo Enterprise Search (CES)는 각 언어의 정립된 단어들을 토큰으로 삼아 분리가 되지 않는 표현으로 인식하고 분리하여 CJKT단어들로 나타났다. 각 CJKT단어는 보통의 단어로 색인된다. 그리하여 CJKT단어의 뜻은 보존이 되며 순위가 한글자 따로따로 매겨지는 것이 아니라 단어로 이뤄지기 때문에 검색능력이 향상된다.

예시: 중국어, 일본어, 한국어, 태국어의 키워드를 검색창에 입력해서 그에 관련된 결과를 찾을 수 있고, 다음 중국어 예시처럼 검색된 CJKT 키워드는 제목과 내용에 하이라이트가 된다.



메모:

- **CES 7.0.6547+ (March 2014)** 향상된 중국어, 일본어, 한국어의 검색능력.
- **CES 7.0.6424+ (February 2014)** 향상된 태국어의 검색능력.
- **CES 7.0.6547+ (March 2014)** 새 색인법으로 새로 향상된 CKJT의 검색능력의 색인법 적용한다. CES를 CES 7.0.6547 에서 CES 7.0.6547+ 의 버전으로 업그레이드시, 기존의 색인법이 기본값으로, 원래의 CJKT 색인법을 사용한다. 새 색인법을 사용하고 싶을 시엔 [Coveo Support](#)에 문의.

아래의 예시에서는 비슷한 대문자의 그룹(예: TTT)은 CJKT단어를 대표하고 다양한 소문자로 이뤄진 그룹(예: abc)은 CJKT가 아닌 단어나 표현을 대표한다.

예시: Coveo supports many languages를 중국어로 단순화해 표현한 것은 다음과 같이 분해된다.

기존 표현: coveo支持多国语言
 대체 표현: abc TTTUUU
 고로: abc는 coveo를 나타냄.
 TTT는 支持 (지원)을 나타냄.
 UUU는 多国语言 (다중언어)를 나타냄.

지원 기능	설명
분류	CJKT의 내용을 각 언어 특유의 유니코드 세트와 인코딩을 기반으로 자동탐색한다. 다른 언어의 단어들과 같이, CJKT의 단어들 또한 어떤 문서의 어느 부분에 나타나는지 감별을 위해 색인된다.
검색	<p>검색시 CJKT표현은 CJKT단어로 분리되어 모든 CJKT의 단어를 포함한 문서의 검색결과로 나타난다. 검색 인터페이스에서는 검색된 CJKT가 결과의 제목과 발체에 하이라이트 된다. 사용자는 CJKT와 비CJKT단어를 혼용할 수 있다.</p> <p>예:</p> <p>입력어: abcTTTUUUdef</p> <p>변환어: (abc TTT UUU def)</p> <p>메모: 결과로 돌아온 문서의 순위는 다른 언어와 같은 방법과 기준을 사용한다.</p>

지원 기능	설명
접두부와 연산자	<p>Coveo 검색 인터페이스의 검색창에 사용자는 CJKT 표현의 접두부와 연산자 사용이 가능하다. 불 연산자는 영어로 쓰여져야 한다 (AND, OR, NEAR, NOT).</p> <p>예시: 사용자는 OR 연산자를 한 단어와 CJKT 표현 사이에 쓸 수 있다:</p> <p>입력어: abc OR TTTUUU 변환어: (abc OR (TTT UUU))</p> <p>예: 사용자는 NEAR 연산자를 한 단어와 CJKT 표현 사이에 쓸 수 있다:</p> <p>입력어: abc NEAR TTTUUU 변환어: (abc NEAR "TTT UUU")</p> <p>입력어: r-cTTTUUU NEAR def 변환어: ("r c TTT UUU" NEAR def)</p> <p>메모: NEAR 연산자는 일치하는 단어나 구절을 지원하나, 하위표현은 지원하지 않는다.</p> <p>예: 사용자는 NOT 연산자를 한 단어와 CJKT 표현 사이에 쓸 수 있다:</p> <p>입력어: NOT TTTUUU 변환어: NOT "TTT UUU"</p> <p>입력어: -TTTUUU 변환어: -"TTT UUU"</p> <p>예: 스테밍은 CJKT에 적용이 되지 않지만, 사용자는 일치의 플러스 (+)나 숫자기호 (#) 연산자를 CJKT표현 앞에 사용하여 정확한 구절로 표현을 늘일 수 있다. 연산자는 지워진다.</p> <p>입력어: +TTTUUU 변환어: "TTT UUU"</p> <p>입력어: #TTTUUU 변환어: "TTT UUU"</p>

지원 기능	설명
구절 검색	<p>Coveo 검색 인터페이스의 검색창에 사용자는 구체적인 CJKT구절을 검색할 수 있다. 구절의 의미는 보존된다.</p> <p>예시: 사용자는 큰따옴표를 사용해 정확히 일치하는 표현으로 한계지를 설정할 수 있다:</p> <p>입력어: abc"TTTUUUdef"</p> <p>변환어: abc "TTT UUU def"</p> <p>단어가 아닌 글자들은 주변 글자들의 정확한 구절을 만들어낸다.</p> <p>예시: 대시 기호의 존재로 (-) 변환이 정확한 구절의 일치가 되게 억압한다:</p> <p>입력어: TTTUUU-VVV</p> <p>변환어: "TTT UUU VVV"</p>
유의어	<p>Coveo 관리자는 CJKT유의어를 입력하여 검색어를 확장할 수 있다.</p> <p>메모: 입력된 CJKT유의어는 CJKT단어에 적용이 되어 CJKT표현과 그의 단어가 동일시 된다.</p> <p>예시: TTTUUU나 TTT UUU가 유의어로 입력이 되어 같은 효과를 갖는다.</p>
필드 검색	<p>태국 표현은 필드 검색에 사용될 수 있다. 필드내에서의 일치는 정확한 구절의 일치의 변환이기 때문에 더 정확하다.</p> <p>예:</p> <p>입력어: @field=abcTTTUUU</p> <p>변환어: @field="abc TTT UUU"</p> <p>입력어: @field=(abc,TTTUUU)</p> <p>변환어: @field=(abc, "TTT UUU")</p> <p>입력어: @field=(abc,"TTTUUU")</p> <p>변환어: @field=(abc, "TTT UUU")</p>
정지단어	<p>A Coveo 관리자는 CJKT단어를 정지단어로 포함시킬 수 있다.</p>
-를 검색하시겠습니까?	<p>단어 교정 어휘(WCL)는 자주 잘못 입력되는 CJKT단어와 표현을 지원하여 -를 검색하시겠습니까? 까를 검색시 제안한다.</p>

메모: wildcard의 사용은 중국어, 일본어, 한국어, 태국어에 지원되지 않는다.