



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-ZH

Publication date: 6/26/2018

Table of Contents

1. 中文日文韩文泰文语言提高了相关性	1
---------------------------	---

1. 中文日文韩文泰文语言提高了相关性

Coveo平台7一直支持中文、日文、韩文和泰文(CJKT)，而现在为这些语言提供了更好的相关。

这些语言不使用空格字符分隔单词。之前，Coveo企业搜索(CES)将每个字符单独索引，就像它是一个单词一样，并使用这些字符对执行检索。这种索引方法能让最终用户查找到内容，但对诸如检索精度和排名而言并非最佳方案。

通过这种新的索引方法，Coveo企业搜索(CES)使用经过验证的语言感知的分词器对表达式进行识别并分解成独立的语义不可分割的字符(以下简称为CJKT单词)，并将每个CJKT单词索引为普通单词。CJKT单词的含义因此得以保留，由于排序是基于单词而不是基于单个字符，从而提高了相关性。

示例：您可以在搜索框中输入中文、日文、韩文或泰文关键词，以获取搜索结果中的相关文档，并在搜索结果标题和摘录中看到突出显示的CJKT关键词，如下中文示例所示。



笔记：

- **CES 7.0.6547+ (March 2014)** 提高了中文、日文和韩文的相关性。
- **CES 7.0.6424+ (February 2014)** 提高了泰文的相关性。
- **CES 7.0.6547+ (March 2014)** 使用新的提高相关性的CJKT索引方法创建新索引。当您从CES 7.0.6547之前的版本升级到CES 7.0.6547+版本时，现有索引默认将继续使用初始CJKT索引方法。如果您想切换到新的CJKT索引方法，请联系Coveo支持部门寻求帮助。

在下面给出的示例中，一组相似的大写字母(例如：TTT)表示CJKT单词，而一组不同的小写字母(例如：abc)表示非CJKT单词或术语。

示例： Coveo支持多国语言的简体中文表达式分解如下：

初始表达式： coveo支持多国语言
 呈现为： abc TTTUUU
 其中： abc 代表coveo
 TTT 代表 支持
 UUU代表多国语言

已支持的特色	说明
索引	基于每种语言特定的Unicode字符集和编码自动检测CJKT内容。与其他语言的单词一样，CJKT单词被编制索引，以识别它们在哪些文档中出现，以及它们在每个文档中出现的位置。
搜索	在检索时，CJKT表达式被拆分为CJKT单词，搜索结果显示包含所有CJKT单词的所有文档。在搜索界面中，搜索到的CJKT单词在搜索结果标题和摘录中高亮显示。最终用户可以搜索与非CJKT单词或术语混合的CJKT单词。 <p>例如：</p> <p>输入检索内容： abcTTTUUUdef</p> <p>转换后检索内容： (abc TTT UUU def)</p> <p>注意：搜索得到的文档使用与其他语言相同的过程和标准进行排序。</p>

已支持的特色	说明
前缀和运算符	<p>在Coveo搜索界面的搜索框中，最终用户可以使用带有CJKT表达式的搜索前缀和运算符。布尔运算符必须用英语拼写。(AND, OR, NEAR, NOT).</p> <p>示例：最终用户可以在单词和CJKT表达式之间使用OR运算符：</p> <p><u>输入检索内容：</u> abc OR TTTUUU</p> <p><u>转换后检索内容：</u> (abc OR (TTT UUU))</p> <p>示例：最终用户可以在单词和CJKT表达式之间使用NEAR运算符：</p> <p><u>输入检索内容：</u> abc NEAR TTTUUU</p> <p><u>转换后检索内容：</u> (abc NEAR "TTT UUU")</p> <p><u>输入检索内容：</u> r-cTTTUUU NEAR def</p> <p><u>转换后检索内容：</u> ("r c TTT UUU" NEAR def)</p> <p>注意：NEAR运算符支持匹配单词或短语，但不支持子表达式。</p> <p>示例：最终用户可以使用NOT或减号运算符，当在CJKT表达式前面时，该运算符将扩展到精确的短语匹配：</p> <p><u>输入检索内容：</u> NOT TTTUUU</p> <p><u>转换后检索内容：</u> NOT "TTT UUU"</p> <p><u>输入检索内容：</u> -TTTUUU</p> <p><u>转换后检索内容：</u> -"TTT UUU"</p> <p>示例：虽然词干技术不适用于CJKT方法，但最终用户仍可以使用CJKT表达式前面的精确匹配加号(+)或数字符号(#)运算符将表达式扩展为精确短语。运算符将被去除。</p> <p><u>输入检索内容：</u> +TTTUUU</p> <p><u>转换后检索内容：</u> "TTT UUU"</p> <p><u>输入检索内容：</u> #TTTUUU</p> <p><u>转换后检索内容：</u> "TTT UUU"</p>

已支持的特色	说明
短语搜索	<p>在Coveo搜索界面的搜索框中，最终用户可以搜索特定的CJKT短语。短语语义将被保留。</p> <p>示例：最终用户可以使用双引号将表达式定界为完全匹配：</p> <p><u>输入检索内容：</u> abc"TTTUUUdef"</p> <p><u>转换后检索内容：</u> abc "TTT UUU def"</p> <p>非单词字符会生成周围字符的精确短语。</p> <p>示例：破折号的存在(-)使其强制转换为精确的短语匹配：</p> <p><u>输入检索内容：</u> TTTUUU-VVV</p> <p><u>转换后检索内容：</u> "TTT UUU VVV"</p>
Thesaurus	<p>每位Coveo管理员都可以在同义词库中输入CJKT表达式以展开查询。</p> <p>注意：CJKT同义词库条目应用于CJKT单词，因此一个CJKT表达式及其CJKT单词被认为是等效的。</p> <p>示例：在同义词词典条目中输入 TTTUUU 或TTT UUU具有相同的效果。</p>
字段检索	<p>泰语表达式可用于字段检索。字段内的匹配更精确，因为它们被转换为精确的短语。</p> <p>示例：</p> <p><u>输入检索内容：</u> @field=abcTTTUUU</p> <p><u>转换后检索内容：</u> @field="abc TTT UUU"</p> <p><u>输入检索内容：</u> @field=(abc,TTTUUU)</p> <p><u>转换后检索内容：</u> @field=(abc, "TTT UUU")</p> <p><u>输入检索内容：</u> @field=(abc,"TTTUUU")</p> <p><u>转换后检索内容：</u> @field=(abc, "TTT UUU")</p>
停止词	<p>Coveo 管理员可以将CJKT单词囊括为停止词。</p>
你在寻找	<p>单词校正器词典(WCL)支持CJKT单词，因此键入错误的CJKT表达式会在查询时指引您前往你在寻找建议。</p>

注意：中文、日文、韩文和泰文不支持使用通配符。