

Tokenizer いろいろ比較

Solr 3.1で利用できるTokenizerを比較しました.

この資料のURL

- http://haruyama.github.com/Solr_5_20110516/
- <http://goo.gl/nG5bG>
 - ネヌジー5ビージー
- <http://bit.ly/ID4K1k>
 - エルディー4ケー1ケー

自己紹介

- 春山 征吾 @haruyama
- 本日(2011/05/16) ECナビを退社.
- 転職先募集中!

- 転職用情報 - 春山征吾のWiki - livedoor Wiki (ウィキ)
- 懇親会やTwitter,
メールなどでご連絡お願い致します.

Tokenizerの比較

データのTSV(4.7GB)をupdate&commitした際の時間とインテ

- [Solr/Tokenizer評価201105](#) - [春山征吾のWiki](#) - [livedoor Wiki \(ウィキ\)](#)

補足

- 1つの大きいTSVを突っ込んだ時の話
 - リソースが許せば,
入力ファイルを分割して並列に更新すれば時間を短縮
 - 私の経験では更新はCPUに束縛される。
ストレージの影響はあるがそれほど多くはない。