

Hachioji.pm #41

# About me

---

## Hideaki Ohno

---

A.K.A. `hide_o_55` / `hideo55`

---

Perl / C++ / Node.js

YAPCトーク応募しています

# 自然言語処理を支える技術 ～要素技術とPerlの活用～

今回もテーマ無視

ウェーブレット行列ライブラリを



作った話

ウェーブレット行列とは

ウェーブレット木の

---

効率的かつ簡易な実装

ウェーブレット木

# 整数列(文字列)に対するデータ構造

**rank():** 文字列の先頭から位置  $k$  までに、文字  $x$  がいくつあるかを返す

`select()`: 文字列の先頭から見て、`n` 個目の文字 `x` の次の位置はどこかを返す

なにがうれしいか

圧縮全文索引を作れる

# FM-Index

---

圧縮全文索引で最速

---

索引からオリジナルを復元できる

---

わかち書きをする必要がない

今回作ったやつ



<https://github.com/hideo55/cpp-HSDS>

# 特徴

---

`wat-array`互換の機能を全て実装

# 特徴

---

高速化した完備辞書

`select`を高速化するための辞書を内部にもつ



SSE4.2(POPCNT)対応



使える場合は組み込み命令を使用

# 特徴

---

mmap対応

次はこれでFM-Index書くか



LOUDS書く予定

以上