

The title 'Hyper Estraier' is centered on the page. It is surrounded by six light purple circles. Three circles are positioned above the text: one behind the 'H', one behind the 'E', and one to the right of the 'r'. Three circles are positioned below the text: one to the left of the 'H', one behind the 'E', and one to the right of the 'r'.

Hyper Estraier

a full-text search system for communities

背景



- 読み方

- ハイパーエストレイ(ア|ヤ)(一)?
- estraier: 迷う、はぐれる = stray

- なぜ作ったか

- 趣味
- Snatcher→QDBM→Estraier→Hyper Estraier

- 何に使うのか

- WebやDeep Webの探索
- SIでの利用(商用)

特徴

- 三拍子揃ってる！
 - はやい = 高性能
 - うまい = 高機能
 - やすい = 簡単
- 「簡単」を最も重視



はやい

- 転置インデックスを使った高速な検索
 - 対象10万件でも0.1秒以内に応答
- インデックスの作成がかなり高速
 - HTMLで4000文書/min.以上の登録速度
- 大規模なサイトでもOK
 - 普通のPCで動作
 - 100万件くらいは余裕



うまい

- N-gram方式による漏れのない検索
 - あらゆる言語を扱える
- 対象文書に依存しない
 - あらゆるリポジトリ、あらゆるフォーマット
 - 任意の属性データ
- 多様な検索方式
 - フレーズ検索、論理式、属性検索、類似検索



やすい

- 導入簡単、高度な応用
 - コマンドツールとCGIスクリプトを提供
 - APIとドキュメントの整備
- どこでも動く
 - Linux、BSD、Mac、そしてWindows
- オープンソース
 - GNU LGPLで配布

デモンストレーション



- インデクシング

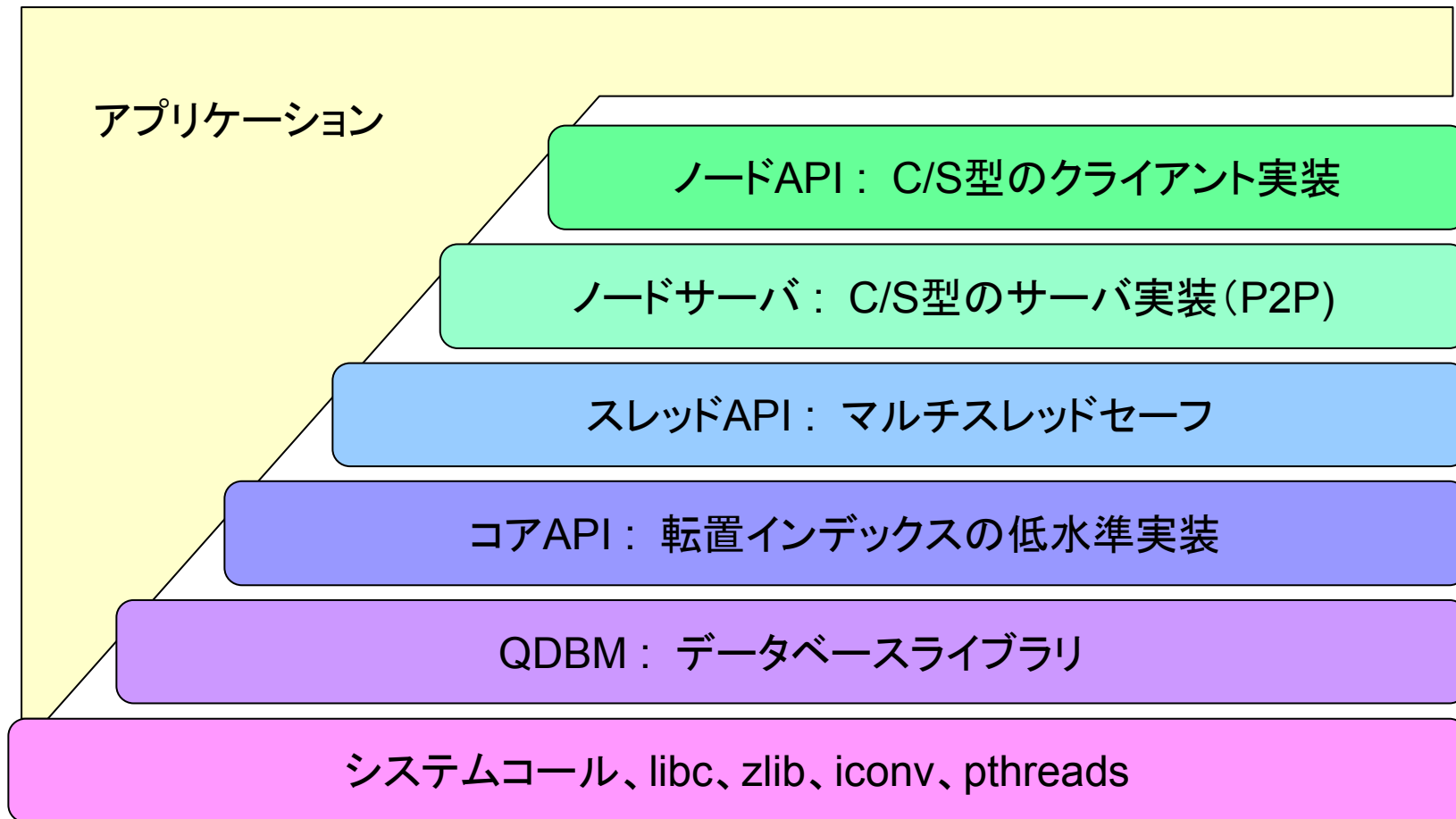
- やたら地味、、、

- 検索

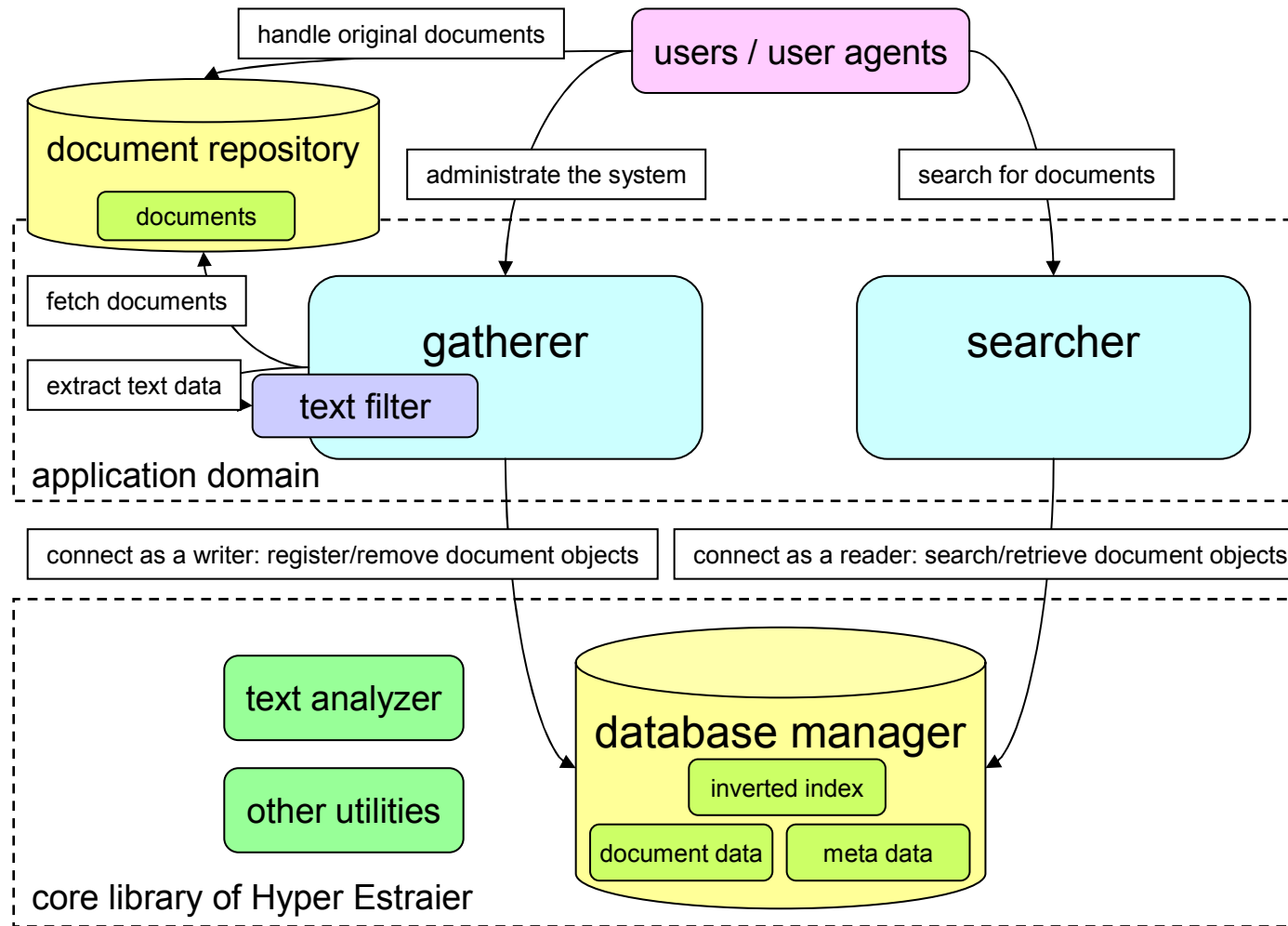
- フレーズ検索、論理式、属性検索、類似検索

- インクリメンタル検索

APIの階層



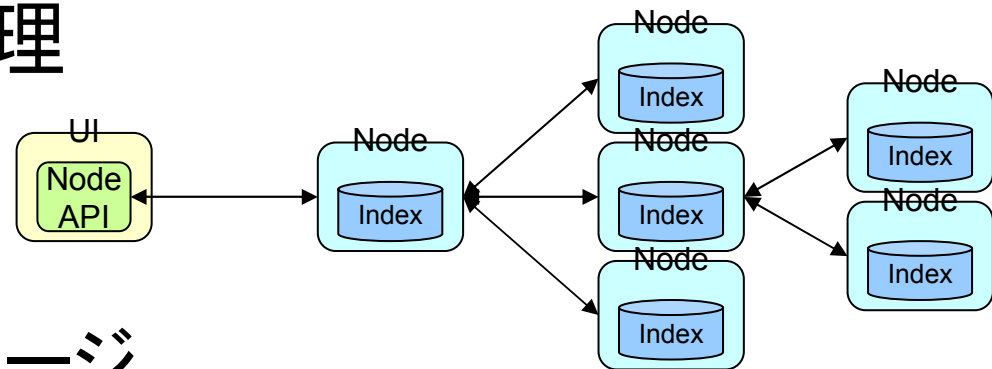
アプリケーションの構成



今後の展望(1)

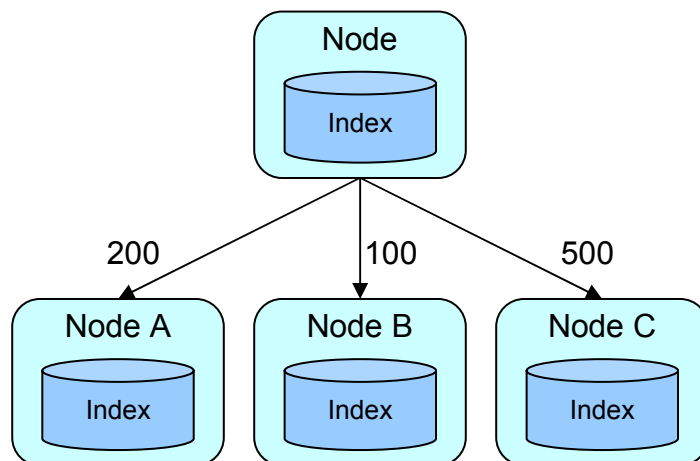
● P2P連携の分散処理

○ 多階層メタ検索



○ 信頼度を加味したマージ

○ クリックフィードバック



各ノードから10件収集した場合のスコア

Node Aの1位 = $10 * 200 = 2000$

Node Aの2位 = $9 * 200 = 1800$

...

Node Bの1位 = $10 * 100 = 1000$

Node Bの2位 = $9 * 100 = 900$

...

Node Cの1位 = $10 * 500 = 5000$

Node Cの2位 = $9 * 500 = 4500$

...

今後の展望(2)

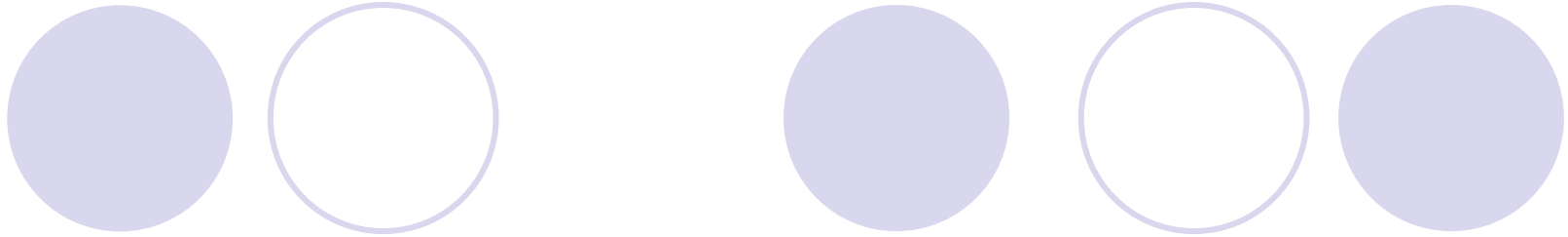


- アプリケーションの拡充
 - 各種言語のバインディング
 - Java、Ruby、Perl
 - mod_estrailer
 - 各種Webアプリケーション連携
 - プロキシ連携、簡易Webサーチエンジン
 - 文書管理システム
 - WebDAV連携
 - 検索ノードファーム



まとめ

- 三拍子揃った全文検索システム
 - その名もHyper Estraier
- P2P連携機能は鋭意実装中
 - 8月末までにはリリース予定
- アプリケーションを作りますか？
 - 厚くサポートする所存
 - アイデアも募集



ありがとうございました。

連絡先:

平林幹雄 <mikio@users.sourceforge.net>

プロジェクトページ:

<http://hyperrestrailer.sourceforge.net/>

メーリングリスト:

<http://lists.sourceforge.net/lists/listinfo/hyperrestrailer-users-ja>