

# CCFinderNexGen(仮称)

--コードクローン検出ツールCCFinderの機能拡張--

神谷年洋  
 科学技術振興機構さきがけ  
 kamiya@mbj.nifty.com

2005/03/15      ソフトウェア工学 第5回セミナー  
 コードクローン検出技術とその応用      1

- CCFinder, Geminiのおさらい
  - 機能
  - 問題点、リクエスト
- CCFinderNexGen(CFNG)の説明
  - 改良点(新機能)
  - リリース予定

2005/03/15      ソフトウェア工学 第5回セミナー  
 コードクローン検出技術とその応用      2

## CCFinder

### コードクローン検出ツール

- 与えられたソースファイルから、コードクローンを検出し、それらの位置を出力する
- CUI、バッチ
- プログラミング言語に応じた変形処理(プリプロセス)
  - + Suffix treeを用いた一致部分文字列検出アルゴリズム
    - コードクローン検出手法に関する特許申請中
- C++, 2万行
  - 現在配布しているのはWindows 2000/XP用のみ
  - gccでもコンパイル可能 (FreeBSD, Mac OS Xで動作確認)

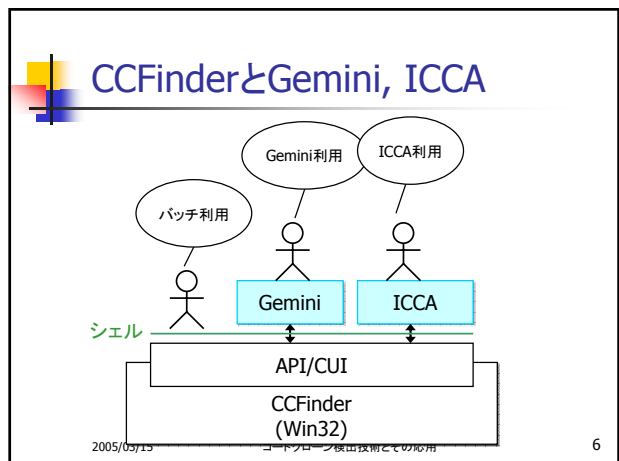
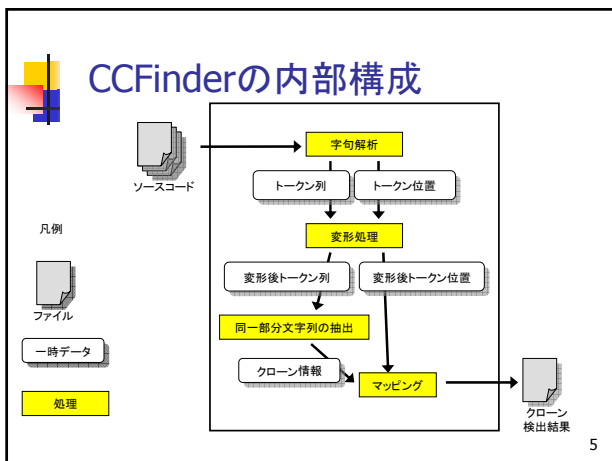
2005/03/15      ソフトウェア工学 第5回セミナー  
 コードクローン検出技術とその応用      3

## Gemini

### コードクローン分析ツール

- コードクローンを視覚的、対話的に分析するツール
- GUI
- 散布図、メトリクスグラフ、ファイル類似度ビュー、ソースコードビュー
- ギャップト・クローン検出機能
  - ギャップト・クローン検出手法に関する特許申請中
- Java、約1万5千行
  - GUIフレームワークとしてSwing(AWT)を利用
  - 内部でCCFinderを利用
  - 開発者: 植田泰士(Jaxa)

2005/03/15      ソフトウェア工学 第5回セミナー  
 コードクローン検出技術とその応用      4



## CCFinderの問題点: (1)ハードコードされた変形処理

- 利用者側でのカスタマイズができない
  - 新しいプログラミング言語に対応するためには、開発者(私)がその文法をきちんと理解して、変形処理ルーチンを開発する必要がある
  - プログラミング言語の方言(独自の文法拡張)に対応できない
  - コメントに書いてある情報を利用することができない
- 異なる文法が混じっているようなソースコードをサポートできない
  - HTMLにJavaが混じっているコード

2005/03/15 ソフトウェア工学 第5回セミナー  
コードクローン検出技術とその応用

7

## CCFinderの問題点: (2)バッチ処理

- インクリメンタルな処理ができない
  - 入力ソースファイルを一つでも修正すると、すべてのソースファイルについて検出し直す必要がある
    - 頻繁なソースコードの修正を前提とする応用で問題
- 微妙な検出条件を設定できない
  - たとえば「特定のソースファイルと類似する部分は除外してコードクローンを検出する」「100トークンから200トークンまでの長さのコードクローンだけを検出する」

2005/03/15 ソフトウェア工学 第5回セミナー  
コードクローン検出技術とその応用

8

## 利用者から頂いた リクエスト (1)

他にもあれば、この機会に是非、ご教示ください

- メモリオプション(-m)の自動的な設定
- 最小一致トークン数オプション(-b)の目安は?
- 変形処理後のトークンの列を見ることができるようになる
- コードクローンの長さを、トークン数ではなく、コメントを除いた行数で
- 特定の処理に関するコードは検出対象から除外したい

2005/03/15 ソフトウェア工学 第5回セミナー  
コードクローン検出技術とその応用

9

## 利用者から頂いた リクエスト (2)

- 大域変数はパラメータ化しないようなオプション
- 自然言語の文書からコードクローンの検出
- 任意のファイル(バイナリファイル)からのコードクローン検出
- あるソフトウェアの2つのバージョンを比較するためのビュー
- Eclipseと連携して動作するように

2005/03/15 ソフトウェア工学 第5回セミナー  
コードクローン検出技術とその応用

10

## 私個人のリクエスト

- 同一のソースコードに対して、異なる検出オプション
  - パラメータ化する・しないの両方
  - ソースファイルの一部に異なる最小一致トークン数を指定する
  - 異なるコードクローン検出アルゴリズムの利用
- Geminiのスケラビリティの改善
  - 理想:CCFinderで扱える→Geminiで扱える

2005/03/15 ソフトウェア工学 第5回セミナー  
コードクローン検出技術とその応用

11

## CNGの開発目標

- 利用を容易にする・カスタマイズを可能にする
  - より多くのプログラミング言語への対応
  - 検出オプション、クエリ、インクリメンタルな検出
- 他のツールとの連携機能を強化する
- 多くのプラットフォームへ対応する

2005/03/15 ソフトウェア工学 第5回セミナー  
コードクローン検出技術とその応用

12

**CNGでの改良点:**

**(1) 字句解析・変形処理の外部化**

- 字句解析と変形処理の手順を、外部定義ファイルで与えられるようにする

→新しいプログラミング言語、方言への対応  
 →異なる文法が混じっているソースファイルへの対応  
 →特定のコードパターンを対象とした変形処理

**CNGでの改良点:**

**(2) 内部データの永続化**

- 以下を永続データとして保持する
  - 検出条件(ソースファイル、検出オプション)
  - 変形処理後のソースファイル、
  - 検出されたコードクローン

→インクリメンタルな検出  
 →入力ソースファイルを後から付け加える  
 →最小一致トークン数を後から小さくする  
 →クエリに基づくコードクローンの問い合わせ  
 →変形処理結果の出力  
 →コメントを除いた行数の出力

**CNGでの改良点:**

**(3) Javaラッパの提供**

- CNGをJavaプログラムからJNI経由で呼び出せるよう、ラッパクラスを作成する
  - Javaプログラムとの緊密な連携

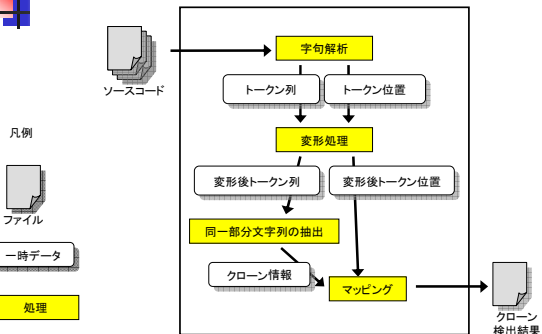
→Mac OS Xへの対応  
 現状で、Mac OS Xではシェル(AppleScript)の違いからCCFinderとGeminiを連携して使えない  
 →Eclipseからの利用  
 →(内部データの永続化とあわせて)Geminiのスケラビリティの改善?

**CNGでの改良点:**

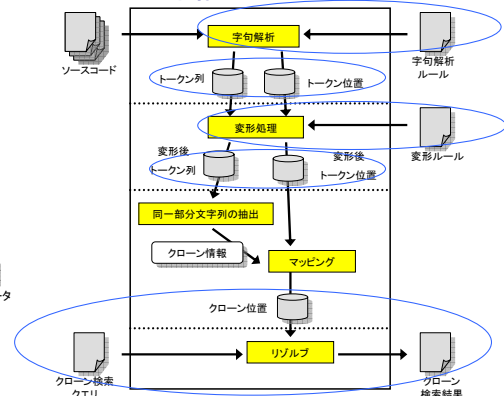
**(4) その他**

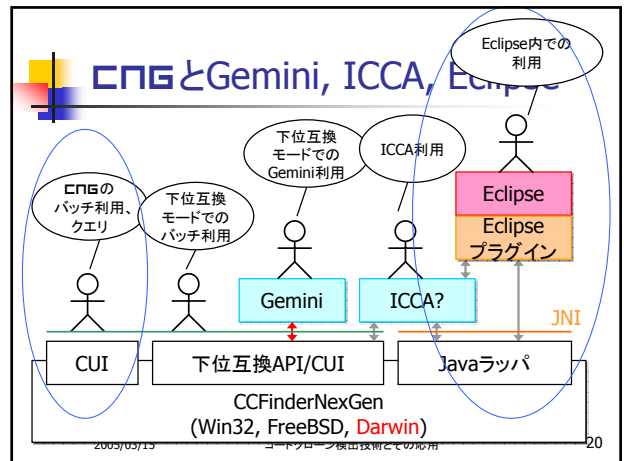
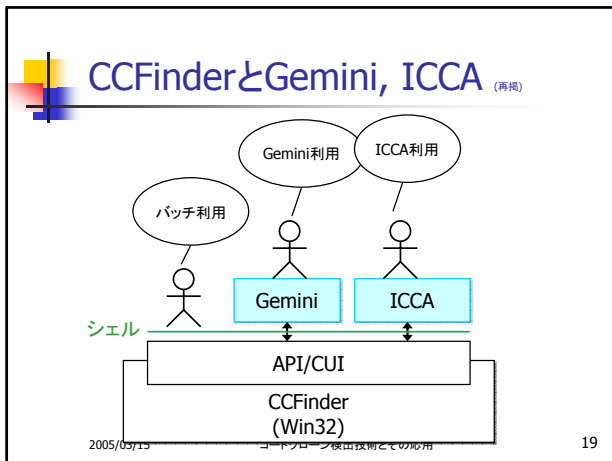
- マルチプル検出(仮称)
  - 同一の入力に対する、検出オプションや検出アルゴリズムの異なる検出
- ギャップ・クローン検出機能をCNGに実装
- 出力フォーマットの変更
  - インクリメンタルな検出、クエリ、マルチプル検出、ギャップ・クローンのために
- ...

**CCFinderの内部構成 (再掲)**



**CNGの内部構成**





- ## 技術詳細
- 変形処理の記述やクエリで用いられるパターンの表現としては、ツリー・オートマトンに基づくものを利用する
    - 括弧の入れ子による構造を素直に表現することができる
      - ←→ 非決定性オートマトン(正規表現)
      - ←→ Perlの動的正規表現
      - ←→ BNF(CFG)
  - 字句解析や変形処理において、外部ツールを利用できるようにする
    - たとえば、日本語の平文を分かち書きするツール
- 2005/03/15      ソフトウェア工学 第5回セミナー  
コードクローン検出技術とその応用      21

- ## リリース予定
- (0) CCFinder、Gemini最終リリース(近日)
    - 以降、バグ修正リリースのみになる予定
    - 下位互換モードによる救済
  - (1) 変形処理の外部的化(8月予定)
  - (2) Javaラッパ(同上)
  - (3) 内部データの永続化(未定)
- (参考) 2004年度第2回未踏ソフトウェア創造事業 梅村 PM「コードクローン検出ツールCCFinderNexGenの開発」
    - <http://www.ipa.go.jp/jinzai/esp/2004mito2/gaiyou/5-7.html>
- 2005/03/15      ソフトウェア工学 第5回セミナー  
コードクローン検出技術とその応用      22

- ## まとめ
- CCFinderの問題を解決するため、CCFGの開発を計画した
    - 利用を容易にする・カスタマイズを可能にする
    - 他のツールとの連携機能を強化する
    - 多くのプラットフォームへ対応する
  - 再設計・再実装が必要で、リリース予定は8月末を予定している
- 2005/03/15      ソフトウェア工学 第5回セミナー  
コードクローン検出技術とその応用      23