

On the Automatic Modularization of Software Systems Using the Bunch Tool

Brian S. Mitchell and Spiros Mancoridis

IEEE Computer Society

22 de abril de 2012

1 Introdução

- Introdução
- Objetivos

2 Software Clustering

- Software Clustering
- Module Dependency Graph
- Abordagem
- Avaliação das Partições do MDG

3 Bunch Tool

- Descrição
- Exemplo
- Estratégia
- Algoritmo de Clusterização
- Outras Features

4 Avaliação

- Estudo de Caso

Introdução

Problemas

- Falta de conhecimento do desenho por parte dos programadores
- Perda da estrutura do sistemas
 - Causando retrabalho ou abandono
- Extração de desenho do código-fonte focam subsistemas isolados
 - Visões apenas das relações de componentes dentro de subsistemas
- Necessidade de compreensão arquitetural do sistema

Objetivos

- Facilitar o entendimento da arquitetura do sistema
- Criação de roteiros para entendimento do software
- Clusterização de subsistemas de um software
 - Utilização de algoritmos de busca
 - Heurísticas
 - Avaliação qualitativa da partição gerada

Software Clustering

- Abordagem independente da linguagem de programação
- Utilização do *Module Dependency Graph* (MDG)
- Clusterização focada nas dependências estáticas do sistema
- Avaliação de qualidade

Software Clustering

Module Dependency Graph

- Representação da estrutura do sistema abstraindo-se a linguagem de programação
- Modela os módulos de um sistema e suas dependências
- Módulo: entidades que encapsulam dados e funções que operam em dados
- Construídas automaticamente através de ferramentas de análise do código-fonte
 - CIA (C), Acacia (C/C++), Chava (Java)

Software Clustering

Abordagem

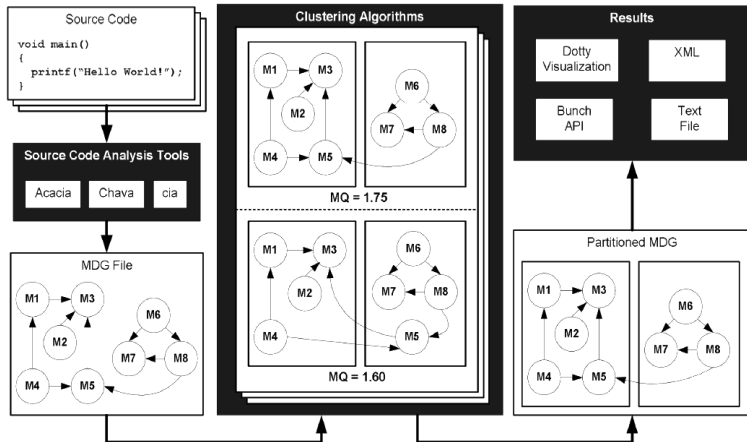
- Particionamento dos módulos do MDG → clusterização
- Busca de um bom particionamento
 - Baixo acoplamento (interdependência)
 - Alta coesão (intradependência)
- Partição: separação dos nós da MDG em conjuntos disjuntos
- Como buscar a solução ótima?
 - Crescimento exponencial do espaço solução
 - Grafo com 15 nós possui 1382958545 possíveis partições!!!

Software Clustering

Avaliação das Partições do MDG

- Impossível analisar todas as possibilidades de partições
- Analogia a um problema de busca
- Maximizar a função avaliadora *Modularization Quality* (MQ)
- MQ: trade-off entre interconectividade e intraconectividade
- MQ valoriza clusters coesos sem alto acoplamento
- Uso de heurísticas
 - Soluções subótimas encontradas rapidamente

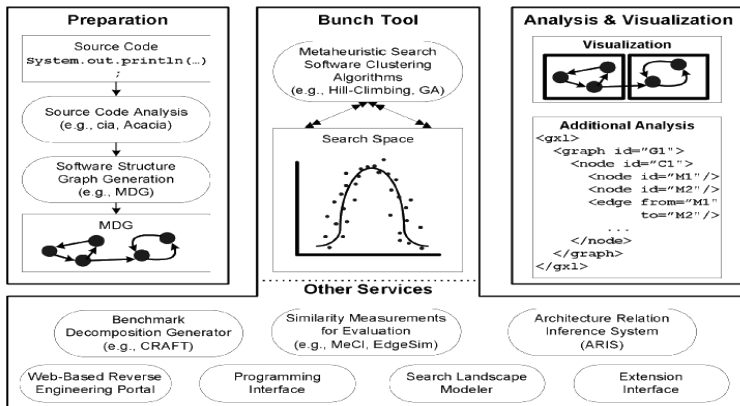
Visão Geral do Algoritmo



Bunch Tool

- Projetada para utilização em projetos de pesquisa
- Flexibilidade para incorporar mudanças
- Portabilidade para diferentes usuários
- Performance
- Extensibilidade para integrar outras ferramentas
- Desenhada mais como framework que aplicação
- Definição da API para promover extensibilidade

Bunch Tool - Arquitetura



Bunch - Exemplo

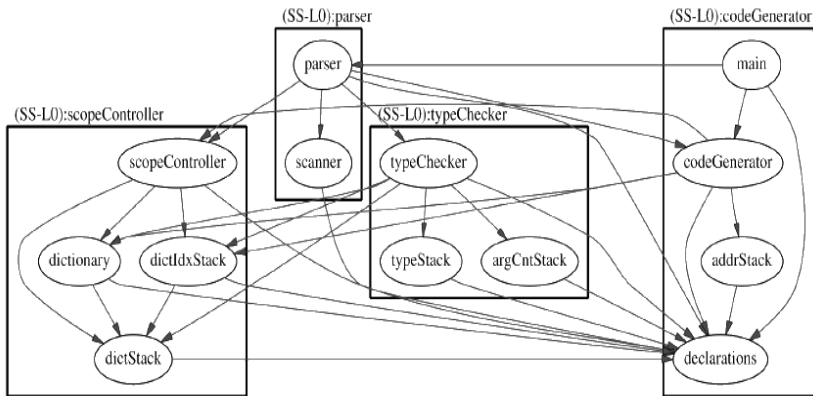
Caso de uso da ferramenta

- Exemplificação feita para um compilador
- Sistema pequeno
- MDG pode ser particionada de 27644437 maneiras

Bunch - Exemplo



Bunch - Exemplo



Bunch - Exemplo

Caso de uso da ferramenta

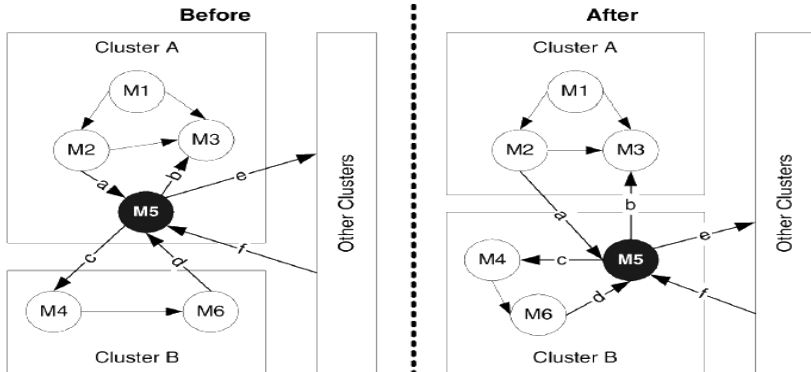
- Uma execução encontrou a solução em 0.07 segundos
- Foram examinadas 373 partições
- Força bruta: 39.57 segundos

Bunch Tool

Avaliação do MQ

- *Modularization Quality* avalia a relação de inter e intradependências
- Soma do *Cluster Factor*(CF), indicador de qualidade por cluster
- CF: peso das arestas intra-cluster / ((peso das arestas inter-cluster) / 2)
 - Penalidade para os dois cluster conectados
- Tentativa de maximizar o MQ através da mudança de módulos por clusters

Bunch Tool



Algoritmo de Clusterização

- Bunch suporta Hill-Climbing e um algoritmo genético
- Hill-Climbing funciona melhor (foco)
- Melhorias levantadas para o algoritmo genético

Hill-Climbing

- Inicia-se com uma partição aleatória do MDG
- Módulos são rearranjados na tentativa de melhorar o MQ
 - Geração de partições vizinhas
- Eventualmente, o algoritmo converge a solução
- Problema: solução encontrada é um máximo local
- Solução: criação de uma população de partições iniciais
- Calibração feita para considerar % de partições vizinhas

Bunch Tool

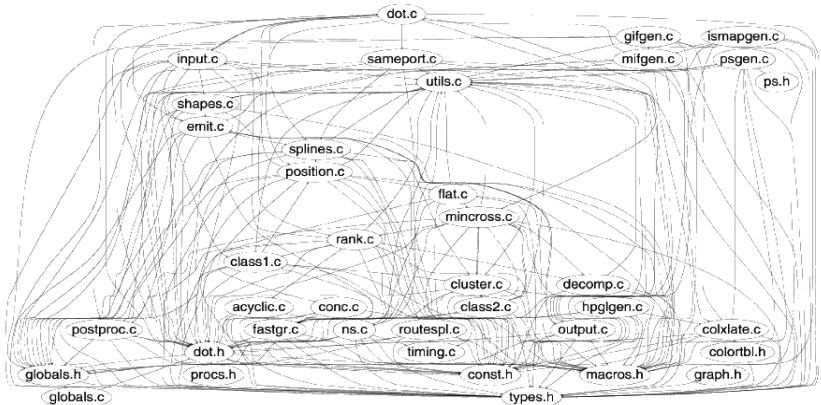
Outras Features

- Remoção de módulos onipresentes
 - Alto número de conexões com outros módulos
- Mostrar/esconder módulos de bibliotecas
 - Detecção de muitas arestas de entradas e nenhuma saída no MDG
- Indicação de clusters pelo usuário

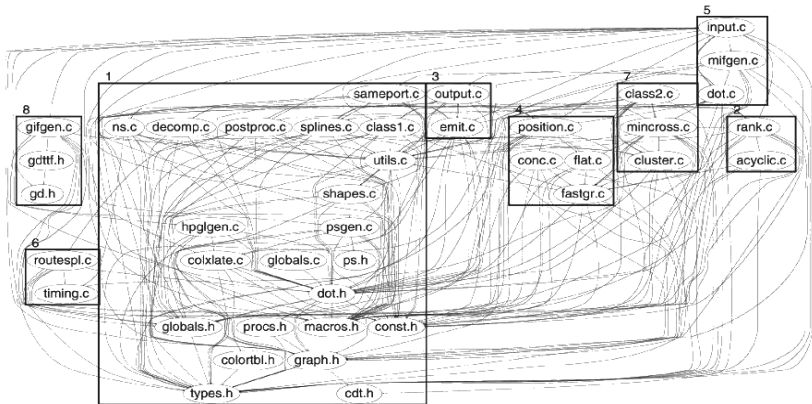
Avaliação

- Feito caso de estudo com a ferramenta de desenho *dot*
- Utilização do Acacia para gerar o MDG
- Particionamento inicial mostrou “anomalias” no sistema
- Presença de módulos onipresentes
 - Arquivos includes com definições comuns ao sistema
- Realocação de módulos onipresentes gerou uma saída pelo Bunch

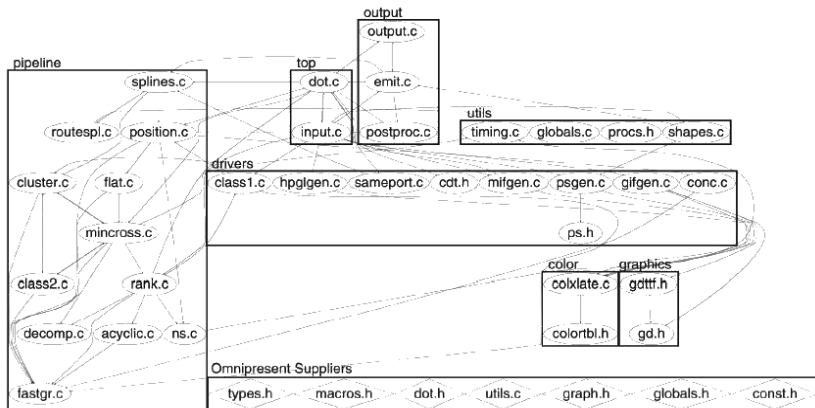
Avaliação



Avaliação



Avaliação



Avaliação Qualitativa

- Feitas diversas análises para outros sistemas com ajuda dos desenvolvedores
- Sistema de arquivos - AT&T
 - Separação entre os módulos do novo sistema de arquivos e a versão antiga
 - Alocação de módulos conectados dentro do mesmo cluster
 - Identificação de módulo que, a princípio, estaria errado.
Análise posterior mostrou que a classificação fazia sentido.

Avaliação Quantitativa

- Necessidade de encontrar uma métrica para avaliação
- Uso de Benchmarks nem sempre é possível
- Criado teste para calcular a similaridade dos agrupamentos feitos por múltiplas execuções
- Se vários resultados contêm o mesmo agrupamento, ele se torna parte da decomposição do sistema
- Grande número de execuções para gerar as estatísticas

Avaliação Quantitativa

- Métricas:
 - EdgeSim: número de intra e inter-arestas que estão presentes em duas partições
 - MeCL: distância (em número de operações de merge) entre duas partições

Métrica	Média	Mínimo	Máximo	Desvio
MeCL	96.5%	92.1%	100.0%	0.17
EdgeSim	93.1%	89.3%	98.6%	0.80

Conclusões

- Ferramenta útil para entender a estrutura do sistema
- Ajuda nas tarefas de manutenção do software
- Algoritmo baseado em heurísticas, encontrando uma solução sub-ótima
- Validação dos resultados necessita da consulta de um expert
- Algoritmos de clusterização podem ser adaptados de acordo com a necessidade
- API desenvolvida para dar suporte a outras ferramentas

OBRIGADO!

gtg@dcc.ufmg.br

lucas.meirelles@dcc.ufmg.br