



# Looking for similarity in source code



<http://www.plade.org/>

Code to analyze

Normalized abstract syntax tree (AST)

**Intra-project**

Source code maintenance and refactoring.  
Revision management.

SortLibrary

QuadraticSortingAlgorithms (QSA)

GnomeSort

```
void gnomeSort(int * a, int n) {
    int i = 0 ;
    while (i < n) {
        if (i == 0 || a[i-1] <= a[i]) i++;
        else {
            int tmp = a[i-1] ;
            a[i-1] = a[i] ; a[i] = tmp ; i-- ; } } }
```

...

BubbleSort

...

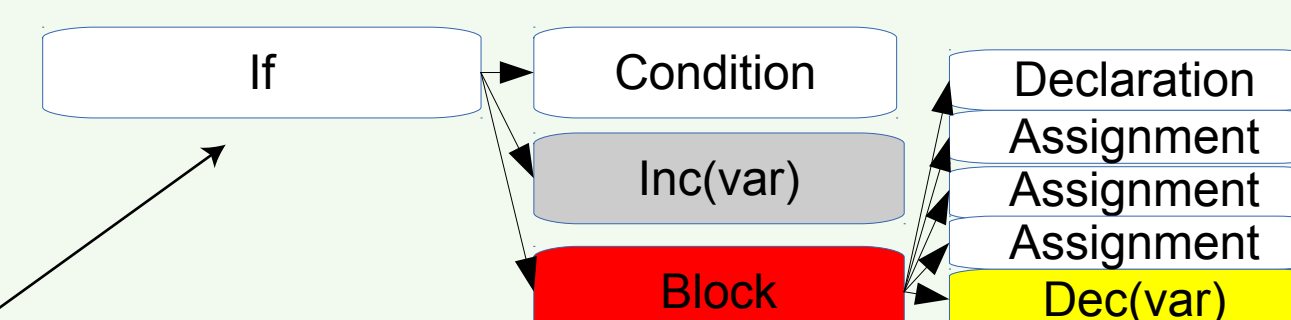
**Inter-projects**

Rightful (or not) source code copying.  
If wrongful : **obfuscation is expected.**

KabouterSort

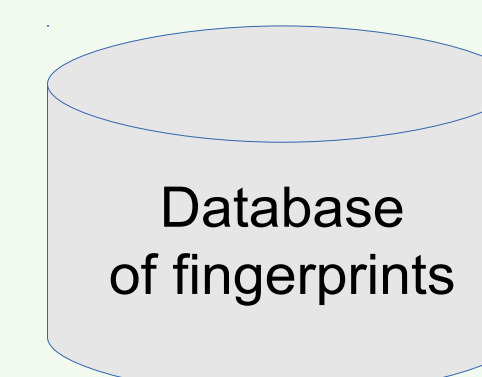
```
void kabouterSort(int m, int b[]) {
    unsigned int j;
    for (j = 1 ; j < n ; ) {
        if (j > 0 && b[j] < b[j-1]) {
            swap(b, j, j-1) ; j-- ; }
        else j = j + 1 ; } }
```

```
void swap(int b[], int i, int j) {
    int tmp = b[i-1/1+0] ;
    b[i] = b[j] ; b[j] = tmp ; }
```



Multi-scale sub-tree fingerprints

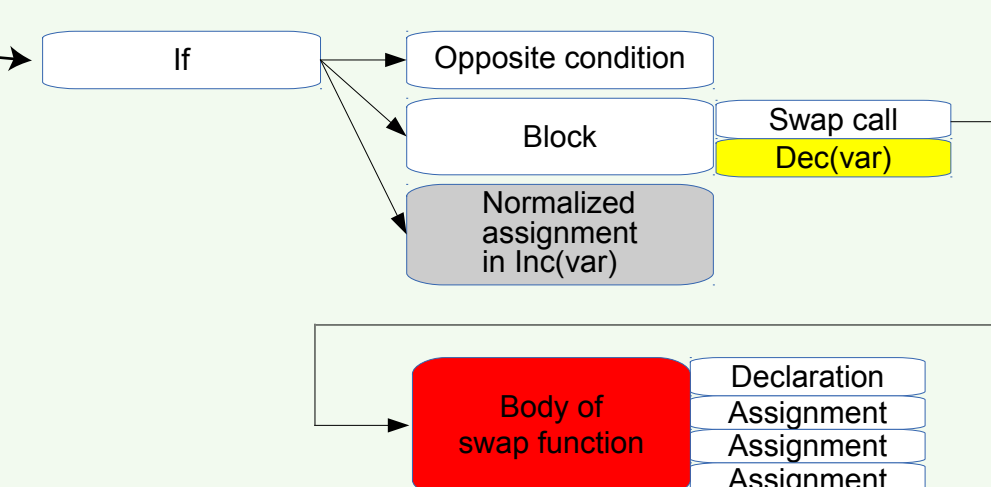
Indexation



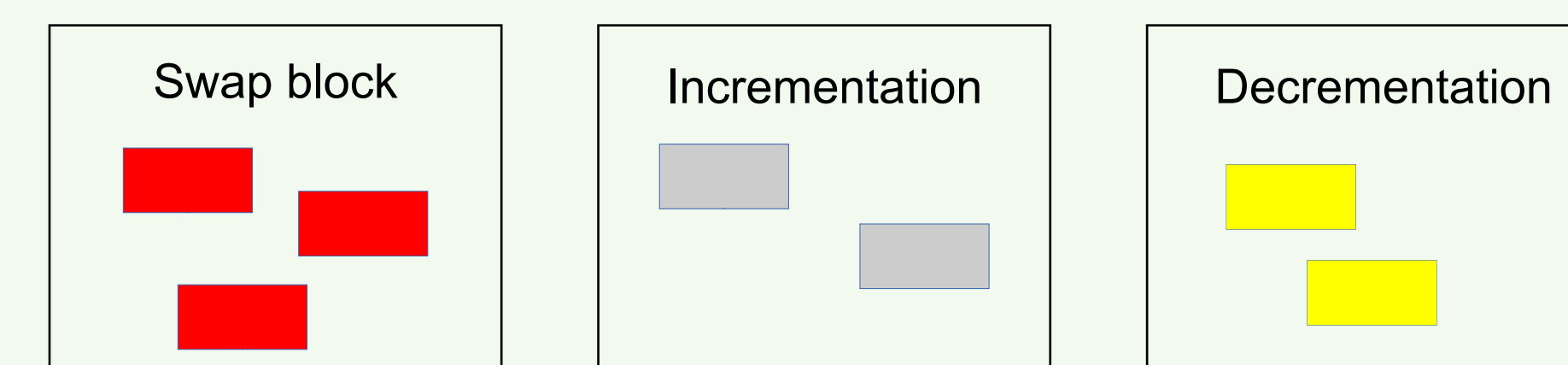
AST normalized



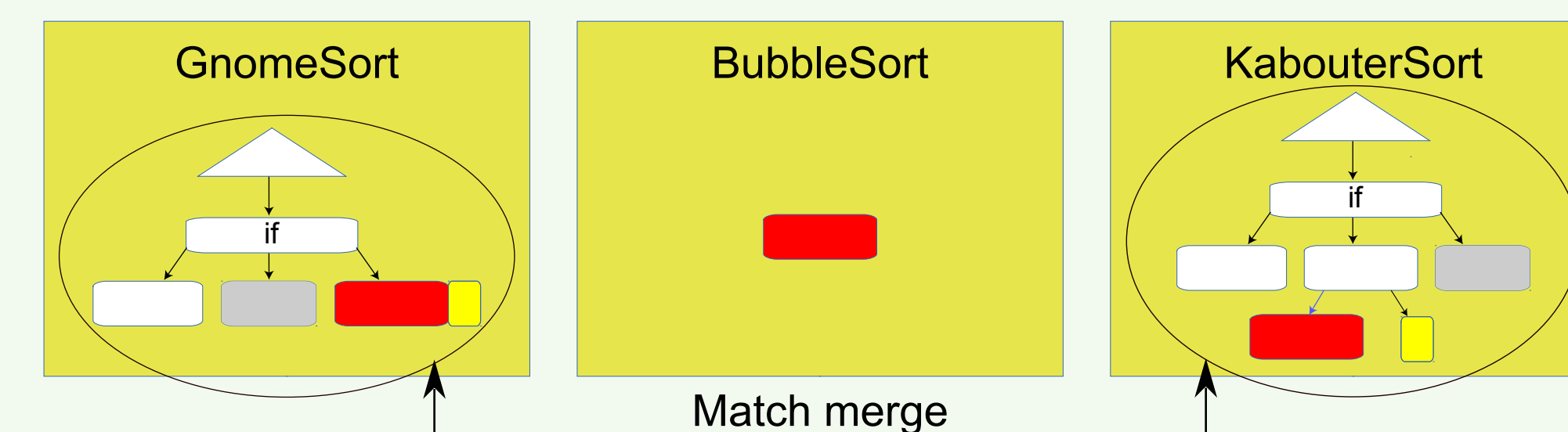
AST normalized



**I - Studying the equivalence classes of the fingerprints**



**II - Consolidating using spatial proximity in the AST**



**III - Analyzing the matches:**  
- quantification and description (scoring)  
- macro-representation (similarity maps, dotplots)

|               | Gnome Sort | Bubble Sort | Kabouter Sort |
|---------------|------------|-------------|---------------|
| Gnome Sort    |            | ~0,2        | ~1            |
| Bubble Sort   | ~0,2       |             | ~0,2          |
| Kabouter Sort | ~1         | ~0,2        |               |

Contact: Michel Chilowicz, Etienne Duris, Gilles Roussel