Figure 1: Pseudocode representation of the integration algorithm

Algorithm: EMAILTOTWITTER

```
1  begin                                                                          prepare
2      pool ← sequence 0 to row length of E;
3      initialize two automated browser sessions via Selenium;
4      while length of pool > 0 do                                                partition
5          if length of pool ⩾ 1000 then
6              batch ← sample of size 1000 from p without replacement;
7          else
8              batch ← pool;
9          remove from batch indices of duplicates in E[, first name] and
             E[, last name];
10         B ← E[batch, ];
11         assign/add B to Blist and write B to disk;
12         pool ← remove from pool the set pool ∩ batch;
13         navigate and log in to Gmail and Twitter in browser sessions;          extract
14         import B from disk in GMail then import contacts in Twitter;
15         T ← extract user data of Twitter contacts via XPath;
16         assign/add T to Tlist;
17         clear Gmail and Twitter contacts;
18         names ← list(split elements in T[, name] into word vectors);           match
19         handles ← list(split elements in T[, handle] into word vectors);
20         cut T[names[every first element] ∩ B[, first name], ] from T and
             integrate in B;
21         cut T[names[every last element] ∩ B[,last name], ] from T and
             integrate in B;
22         cut T[names[every second element] ∩ B[,first name], ] from T
             and integrate in B;
23         cut T[names[every second element] ∩ B[,last name], ] from T
             and integrate in B;
24         cut T[handles[every last element] ∩ B[,last name], ] from T and
             integrate in B;
25         assign/add B to Mlist
26     return list(Mlist, Blist, Tlist)
27 end
```