Input: List of arms A, exploration parameter c
Initialize $t \leftarrow 0$;
Initialize $N(a) \leftarrow 0$ for all arms $a \in A$;
Initialize $R(a) \leftarrow 0$ for all arms $a \in A$;
while $t < T$ **do**
    Increment $t \leftarrow t + 1$;
    Compute the upper confidence bounds:
    $UCB1(a) \leftarrow R(a) + c\sqrt{\frac{\log(t)}{N(a)+\epsilon}}$ for all arms $a \in A$;
    Select the arm with the highest upper confidence bound:
    $a_t \leftarrow \arg\max_{a \in A} UCB1(a)$;
    Pull arm a_t and observe the reward r_t;
    Increment the count of arm a_t: $N(a_t) \leftarrow N(a_t) + 1$;
    Update the estimated reward of arm a_t: $R(a_t) \leftarrow R(a_t) + \frac{r_t - R(a_t)}{N(a_t)}$;
end