ACLOSE algorithm for CFI Mining

1. A frequent itemset - set of items occurring a certain percentage of the time.
2. A closed itemset is set of items which is as large as it can possibly be without losing any transactions.
3. A maximal frequent itemset is a frequent itemset which is not contained in another frequent itemset. Unlike closed itemsets, maximal itemsets do not imply anything about transactions.

Two Functions to be used in Aclose:

t(x) and i(y). The function t(x), where $x \subseteq I$, I being the set of all items, yields the set of transactions that contain x.

Similarly, i(y), where $y \subseteq T$, T being the set of all transactions, yields the set of items that are contained in every transaction of y

transaction ID	items
1	{A,C,T,W}
2	{C,D,W}
3	{A,C,T,W}
4	{A,C,D,W}
5	{A,C,D,T,W}
6	{C,D,T}

Example 1: $t(\{C\})$ The itemset $\{C\}$ is found in all transactions, so the result is $t(\{C\}) = \{1, 2, 3, 4, 5, 6\}$. Taking this result and calling i on it produces the original input set (i.e., $i(\{1, 2, 3, 4, 5, 6\}) = \{C\}$).

Example 2: $t(\{C, D\})$ The itemset $\{C, D\}$ is found in four transactions (i.e., $t(\{C, D\}) = \{2, 4, 5, 6\}$). Similarly to the first example, calling i on this result produces the original input set (i.e., $i(\{2, 4, 5, 6\}) = \{C, D\}$).

Example 3: $t(\{A\})$ The itemset $\{A\}$ is found in four transactions (i.e., $t(\{A\}) = \{1, 3, 4, 5\}$). However, unlike the first two examples, calling i **does not** produce the original set, as $i(\{1, 3, 4, 5\}) = \{A, C, W\}$.

- only two things can happen when calling i and t functions successively on a subset, x
- 1. The subset remains the same (e.g., $i(t(x)) = x$).
- 2. The subset increases slightly, but will never increase again thereafter (e.g., if $i(t(x)) = y$ and $x \subset y$, then $i(t(y)) = y$).
- $f(x) = i(t(x))$ and $g(x) = t(i(y))$. These two new functions are called closure operators.
- Idempotence For all inputs, if the operator is applied twice, it must produce the same result as if it were applied once (e.g., $f(f(x)) = f(x)$).
- Extension The original input subset must be contained in the resulting subset (e.g., $x \subseteq f(x)$).
- Monotonicity If one input subset is contained in another input subset, then the first resulting subset must be contained in the second resulting subset (e.g., if $x \subseteq y$ then $f(x) \subseteq f(y)$).

- Closed Set A set, x, is considered closed w.r.t. closure operator f if it satisfies the property $f(x) = x$, i.e., its closure is itself. Conversely, a set is considered not closed if $x \subset f(x)$ (i.e., if the set gains elements when calling $f(x)$). Note that $x \subseteq f(x)$ holds always.
- {C} is closed since f({C}) = i(t({C})) = {C}. However, {A} is not closed, as f({A}) = i(t({A})) = {A, C, W}. But {A, C, W} is closed
- Generator A set x is called a generator of y if the closure of x is y.
- {A} and {A, C, W} are generators of {A, C, W}.
- Minimal Generator A set, x, is called a minimal generator of y if it satisfies two properties:
- 1. The closure of x is y.
- 2. No proper subset of x generates y.
- {A} is a minimal generator of {A, C, W}. However, {A, C, W} is not a minimal generator as {A} is a subset of {A, C, W} which is itself a generator.

transaction ID	items
1	{A,C,D}
2	{B,C,E}
3	{A,B,C,E}
4	{B,E}
5	{A,B,C,E}

Level 1:

sets	support
{A}	3
{B}	4
{C}	4
{D}	1
{E}	4

Since $\{D\}$ has a support of only 1, we can eliminate it from the search. Applying the *AS* strategy to the remaining four elements gives us the following subsets for level two:

Level 2:

sets	support
{A,B}	2
{A,C}	3
{A,E}	2
{B,C}	3
{B,E}	4
{C,E}	3

- {A, B} and {A, E} can be eliminated, as they both have insufficient support.
- remove sets {A, C} and {B, E}. This is because both have the same support as one of their subsets (e.g., {A, C} has a support of 3 and so does {A}). If a set has the same support as as one of its subsets, it cannot be a minimal generator
- only two remaining sets, {B, C} and {C, E}
- Deriving Closed Sets The second step of the A-Close algorithm involves taking the generators found in the first step and inputting them into f(x) to obtain the closed sets

Level 2:

generators	$i(generators)$	$t(i(generators))$
{A}	{1,3,5}	{A,C}
{B}	{2,3,4,5}	{B,E}
{C}	{1,2,3,5}	{C}
{E}	{2,3,4,5}	{B,E}
{B,C}	{2,3,5}	{B,C,E}
{C,E}	{2,3,5}	{B,C,E}

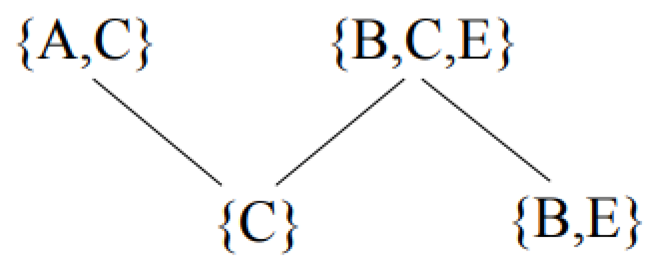

Figure 2: The relationships between the discovered closed itemsets.

After removing duplicates, we find four closed sets: $\{A, C\}$, $\{B, E\}$, $\{B, C, E\}$, and $\{C\}$.