

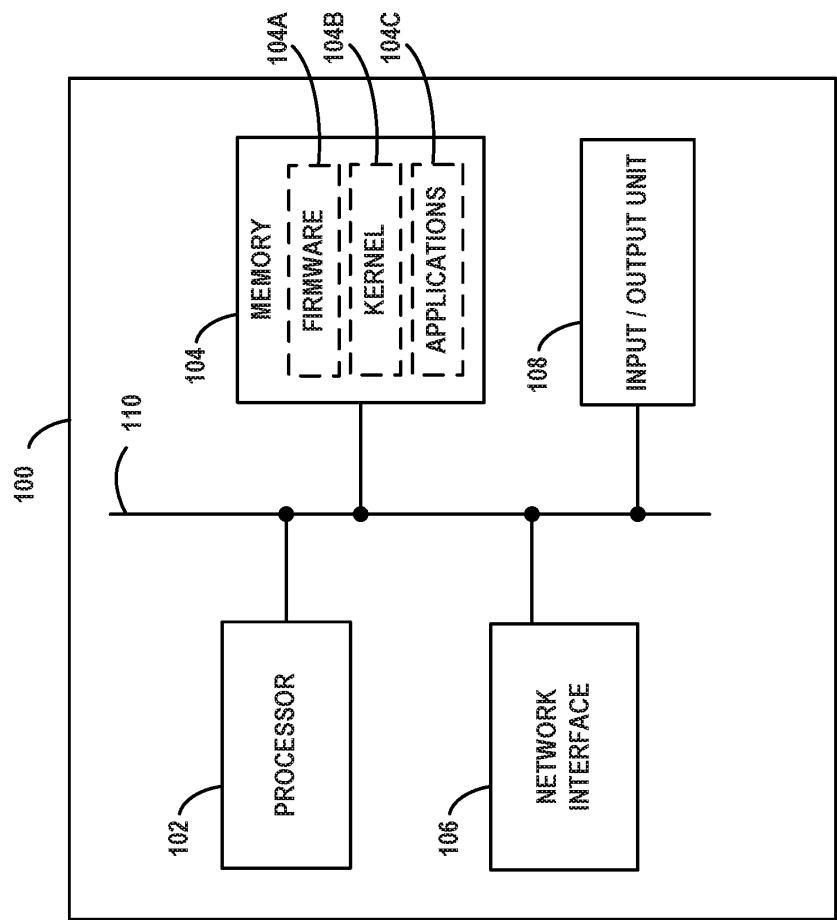
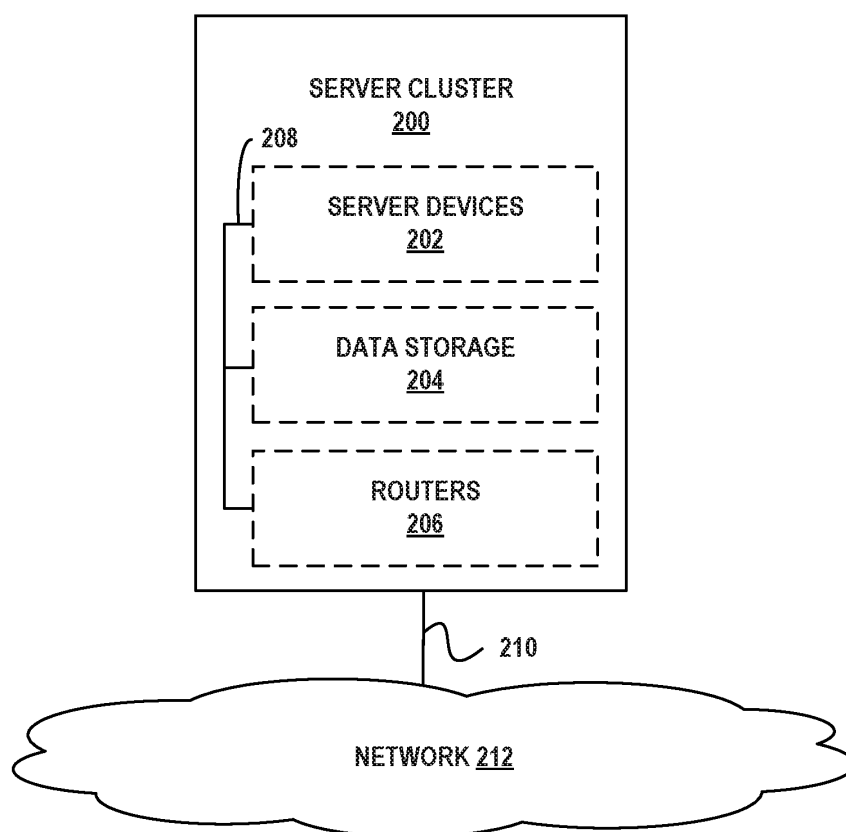


FIG. 1

**FIG. 2**

← 300

1. **make_node** (data_set) { /* data set is the current set of observations */
2. Node current_node /* This is where we store information about the new node that we are adding to the tree */
- 3.
4. **calculate** the entropy of the labels in data_set
5. **for each** of input parameters in data_set:
6. **calculate** the information gain with respect to the labels
7. **select** the input parameter with the highest information gain
8. **designate** the selected input parameter for current_node
9. **remove** the selected parameter from data_set
10. **for each** value v of the selected input parameter:
11. **create** edge for v descending from current_node
12. **if** (the label has a value of "yes" for all observations in data_set in which the selected parameter has a value of v):
13. **return** (current_node as a leaf node designated "yes" connected to edge)
14. **if** (the label has a value of "no" for all observations in data_set in which the selected parameter has a value of v):
15. **return** (current_node as a leaf node designated "no" connected to edge)
16. **make_node** (data_set)

FIG. 3

4/24

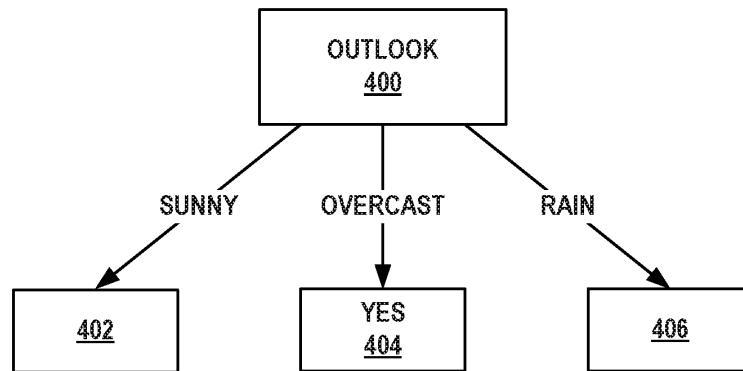


FIG. 4A

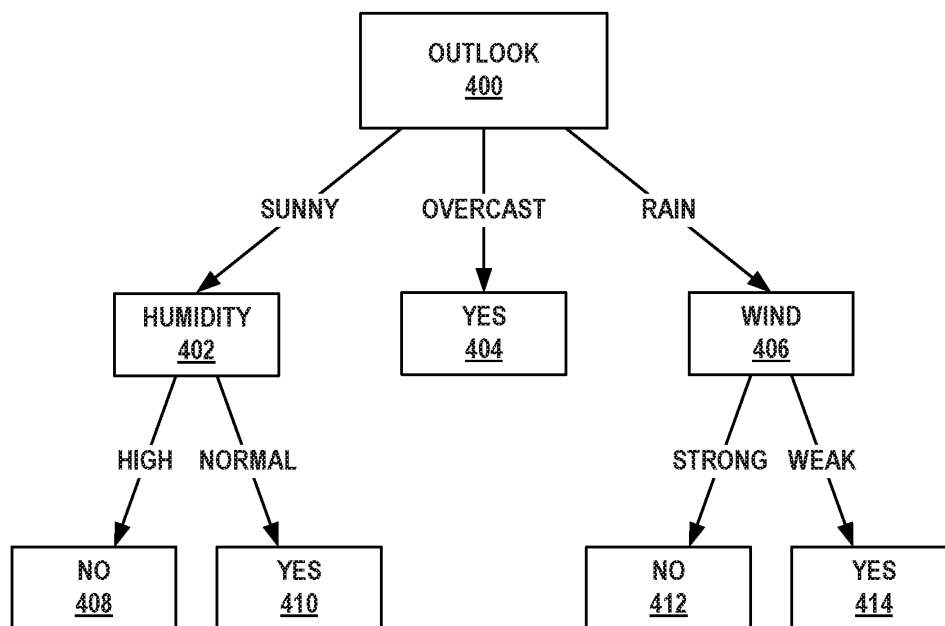


FIG. 4B

- 39 Plastics and articles thereof
- 3901 Polymers of ethylene, in primary forms:
- 3902 Polymers of propylene or of other olefins, in primary forms:
- 3903 Polymers of styrene, in primary forms:
- 3904 Polymers of vinyl chloride or of other halogenated olefins, in primary forms:
- 3905 Polymers of vinyl acetate or of other vinyl esters, in primary forms; other vinyl poly
- 3906 Acrylic polymers in primary forms:
- 3907 Polyacetals, other polyethers and epoxide resins, in primary forms; polycarbonate:
- 3908 Polyamides in primary forms:
- 3909 Amino-resins, phenolic resins and polyurethanes, in primary forms:
- 3910000000 Silicones in primary forms
- 3911 Petroleum resins, coumarone-indene resins, polyterpenes, polysulfides, polysulfor
- 3912 Cellulose and its chemical derivatives, not elsewhere specified or included, in prim
- 3913 Natural polymers (for example, alginic acid) and modified natural polymers (for ex
- 3914 Ion-exchangers based on polymers of headings 3901 to 3913, in primary forms:

FIG. 5

6/24

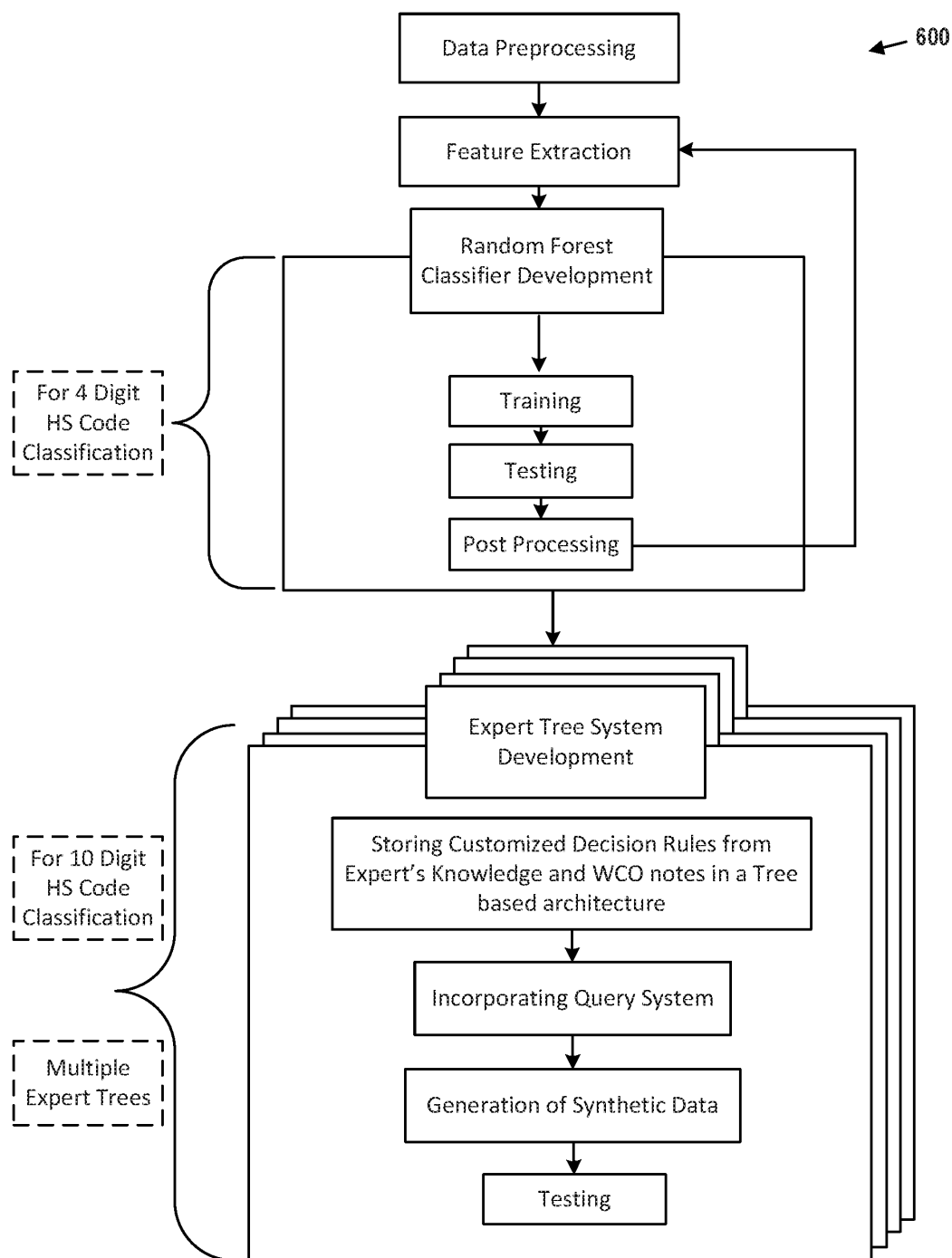


FIG. 6

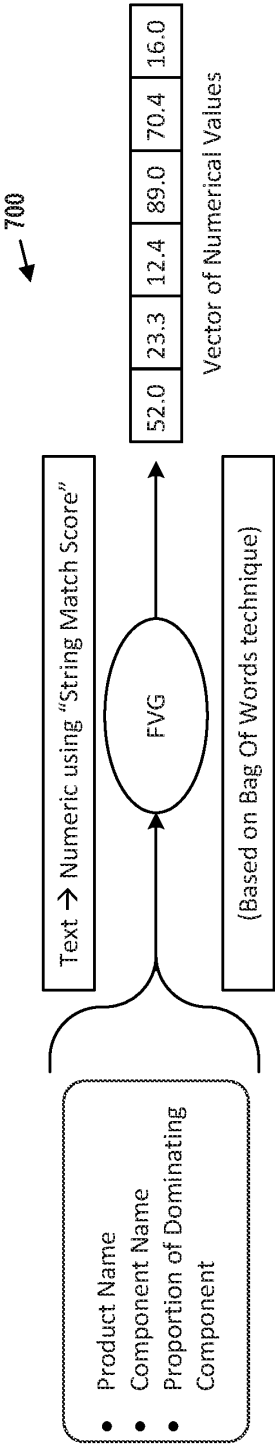


FIG. 7

- 
- Step 1: Merge product names of all the products with 3901 as the first 4 digits of their HS-Code and store them in a file T_p .
- Step 2: Remove punctuation and stopwords (set of words which don't carry useful information, such as: of, with, the, a, on, etc) and update T_p .
- Step 3: Expand the acronyms and update T_p .
- Step 4: Count the frequency of each words in T_p .
- Step 5: Set the threshold as $t=0.5 \times F_{avg}$, where F_{avg} is the average of all frequency counts computed in the previous step.
- Step 6: Select the words with their frequency count larger than the threshold value t and store them in a separate file BOW_p .
- Step 7: This BOW_p is the P-File for HS-Code 3901.

FIG. 8A

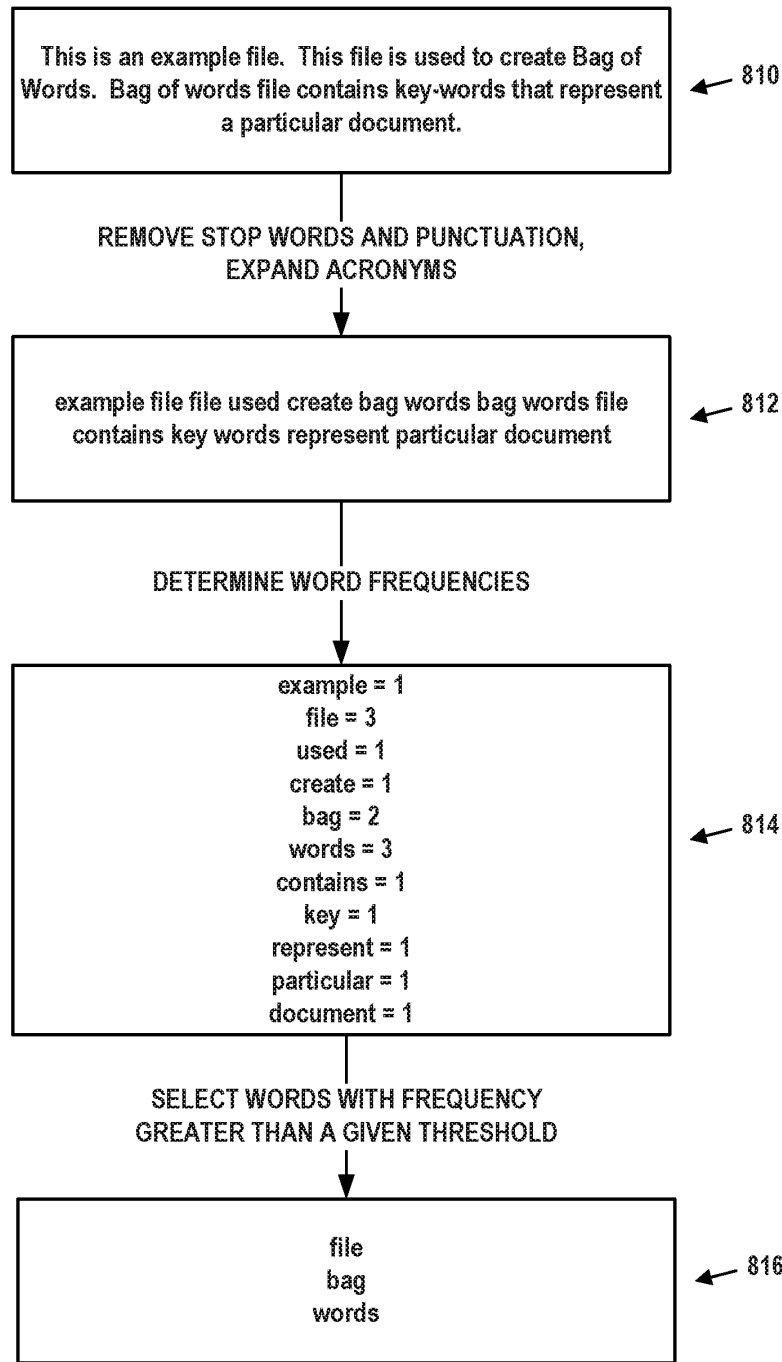


FIG. 8B

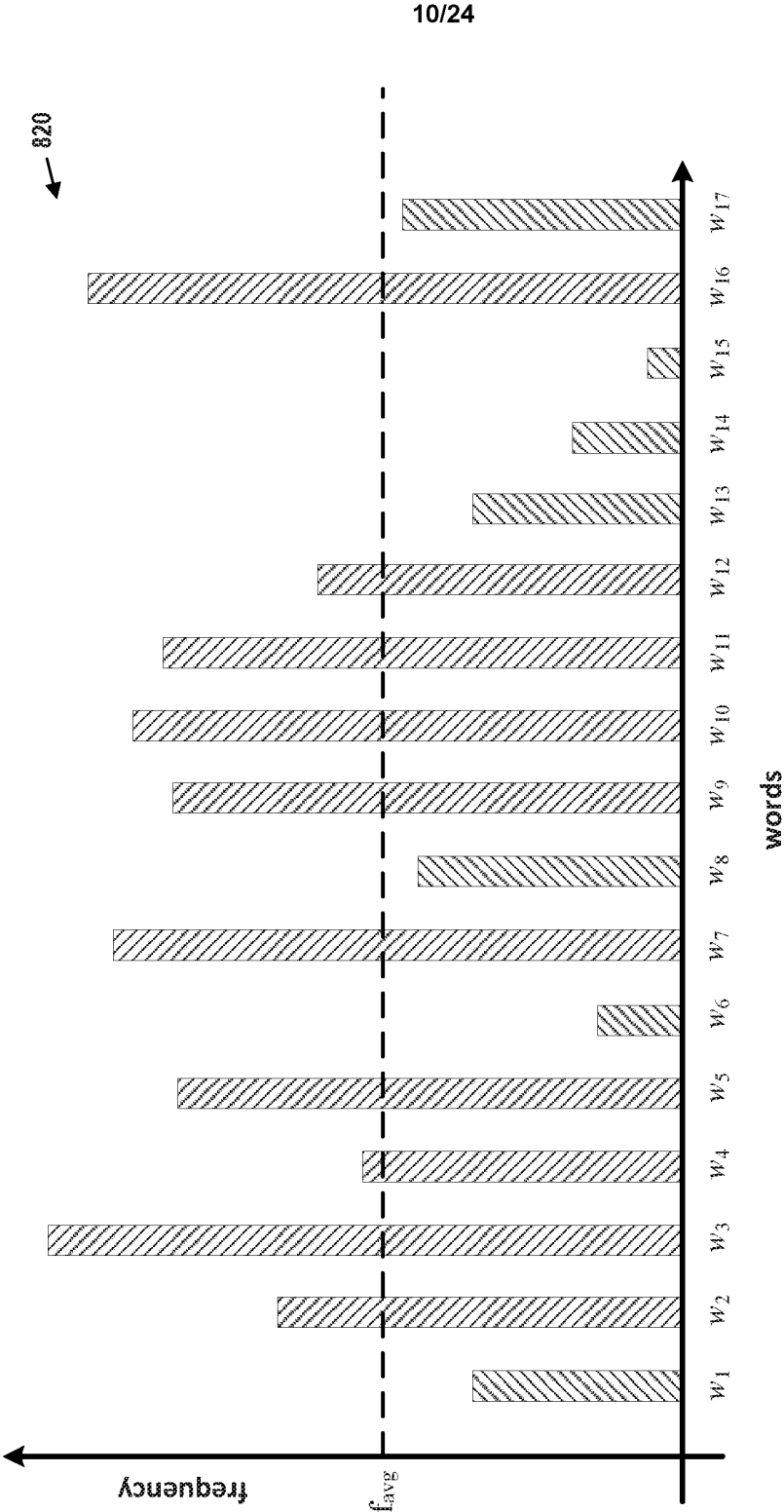


FIG. 8C

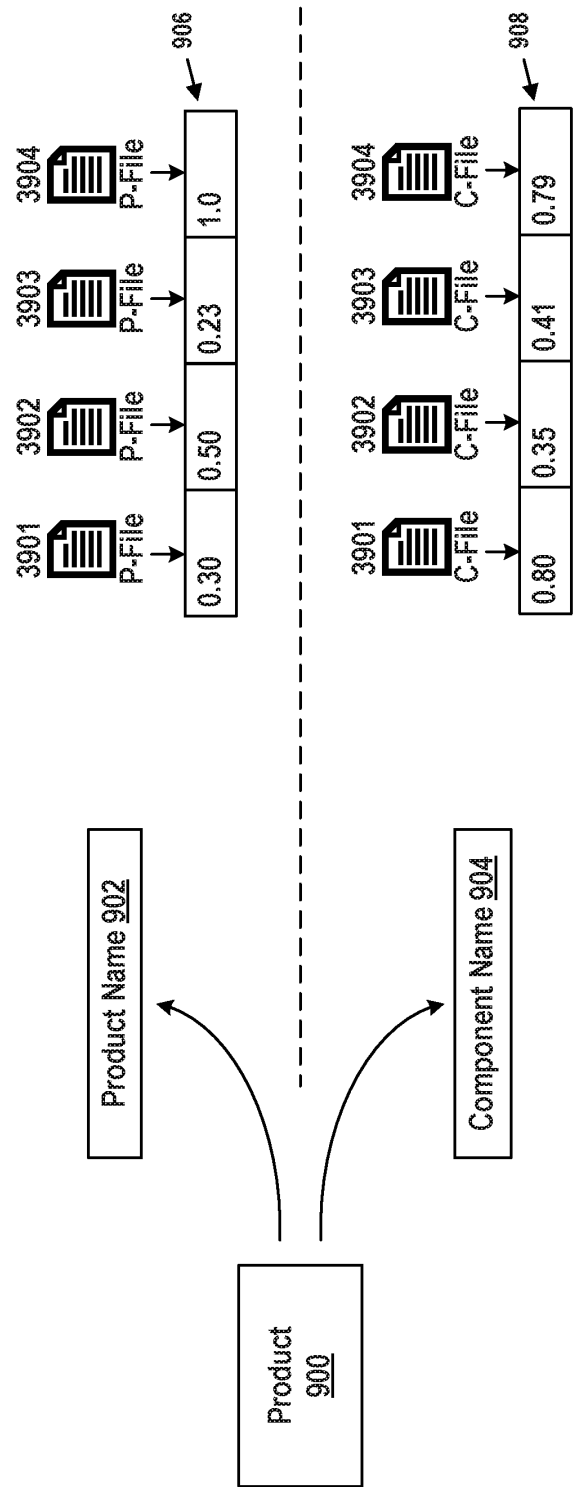


FIG. 9

12/24

1000
←

```
1. Preprocessing: preprocess the product name of and tokenize it to obtain a set of words  $S_w$ 
2. Initialize: net-match-score = 0
3. for w in  $S_w$  do
4.   best-match-score = 0;
5.   for each word  $w_b$  in an existing P-file do
6.     match-score = 1 - normalizedLevenshtein(w,  $w_b$ );
7.     if match-score > best-match-score then
8.       best-match-score = match-score
9.     end
10.  end
11.  net-match-score += best-match-score
12. end
```

FIG. 10A

13/24

```

1. Preprocessing: preprocess the product name and tokenize it to obtain a set of words  $S_w$ 
2. Initialize: net-match-score = 0
3. for w in  $S_w$  do
4.   best-match-score = 0;
5.   best-weight = 1;
6.   for each word  $w_i$  in an existing P-file do
7.     match-score = 1 - normalizedLevenshtein( $w, w_i$ );
8.     if match-score > best-match-score then
9.       best-match-score = match-score;
10.      best-weight = weight( $w_i$ );
11.      //where weight( $w_i$ ) is the weight of word  $w_i$  in an existing P-file
12.    end
13.  end
14.  net-match-score += (best-weight * best-match-score)
15.
16. for each RegEx R in an existing P-file do
17.   pattern-match-score = compare( $w, R$ );
18.   //pattern-match-score can be 0 (no match) or 1 (100% match)
19.   net-match-score += (1000 * pattern-match-score)
20. end
21. end

```

Algorithm to compute the net-match-score (or string-similarity score) of the product name and the P-file with weighted list of words and regular expressions.

FIG. 10B

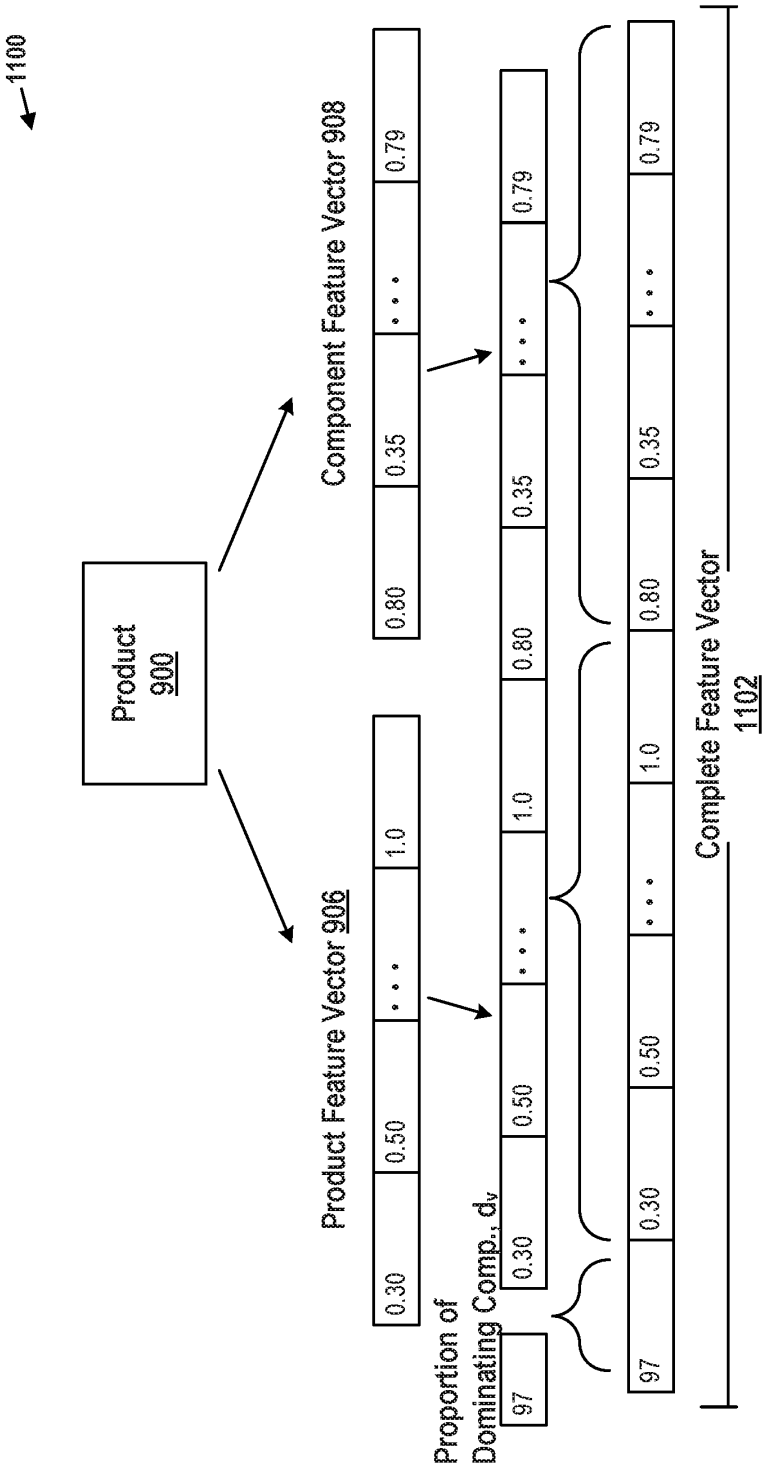
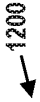


FIG. 11

Step 1: For each word in a P-file or C-file, determine its frequency count 

Step 2: Identify $f_{\min}$ and $f_{\max}$ as minimum and maximum frequency counts within a file

Step 3: Compute the weight (w) of each word using the following equation:

$$w = 1 + (f_w - f_{\min}) / (f_{\max} - f_{\min})$$

This will ensure that weight of each word will lie between 1 and 10. The weight indicates the significance level of the word, with the value of 1 indicating lowest significance level and the value of 10 indicating the highest significance level.

Step 4: Keywords suggested by experts are given weight-value of 5, 7 or 10 as per an expert's comment about their relevance level



chlorinated::	1.03765690377
acetic::	1.15062761506
butene::	1.94142259414
ethylene::	10.0
hexene::	4.05020920502
ester::	1.37656903766
acrylic::	1.0
octene::	5.66945606695
ethene::	9.58577405858
ethenyl::	1.22594142259

FIG. 12A

FIG. 12B

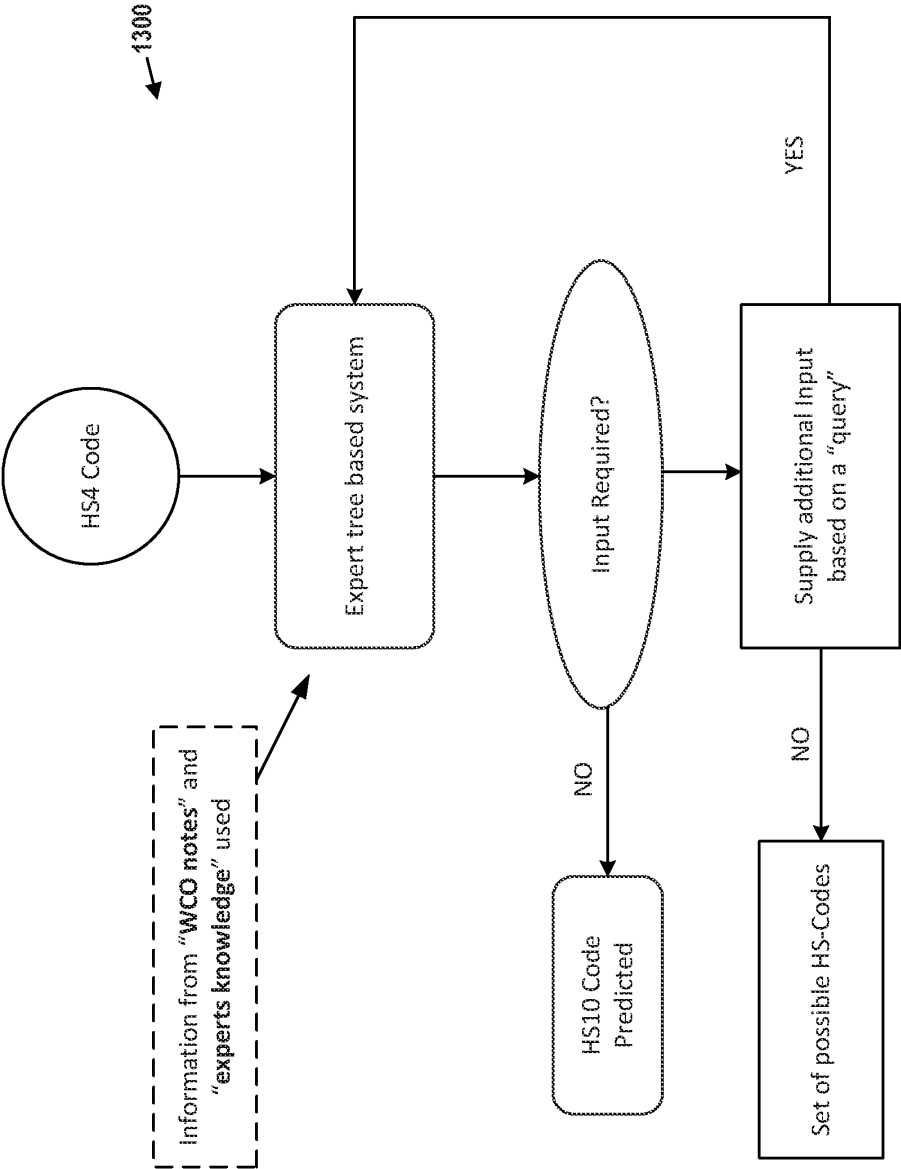


FIG. 13

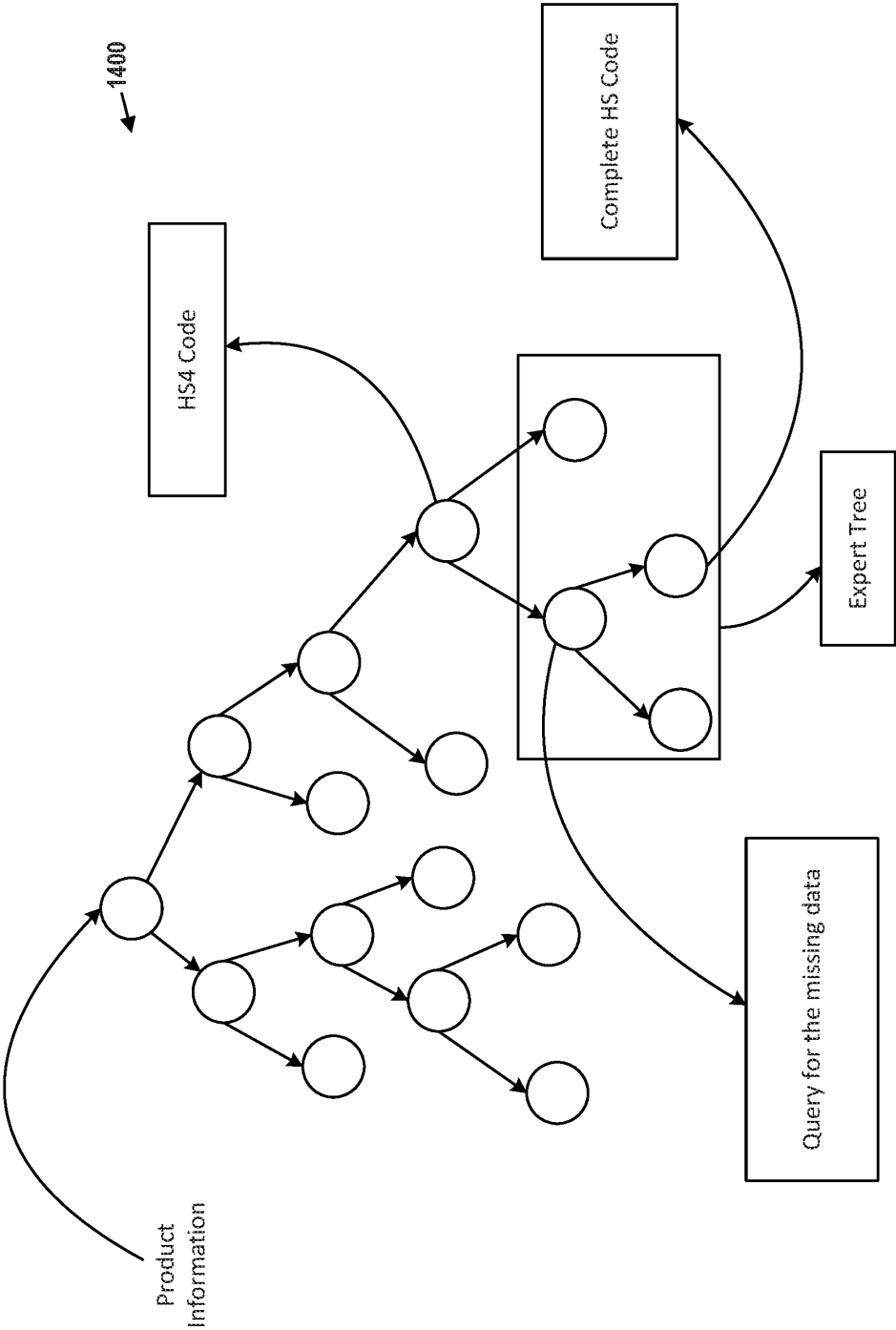


FIG. 14

18/24

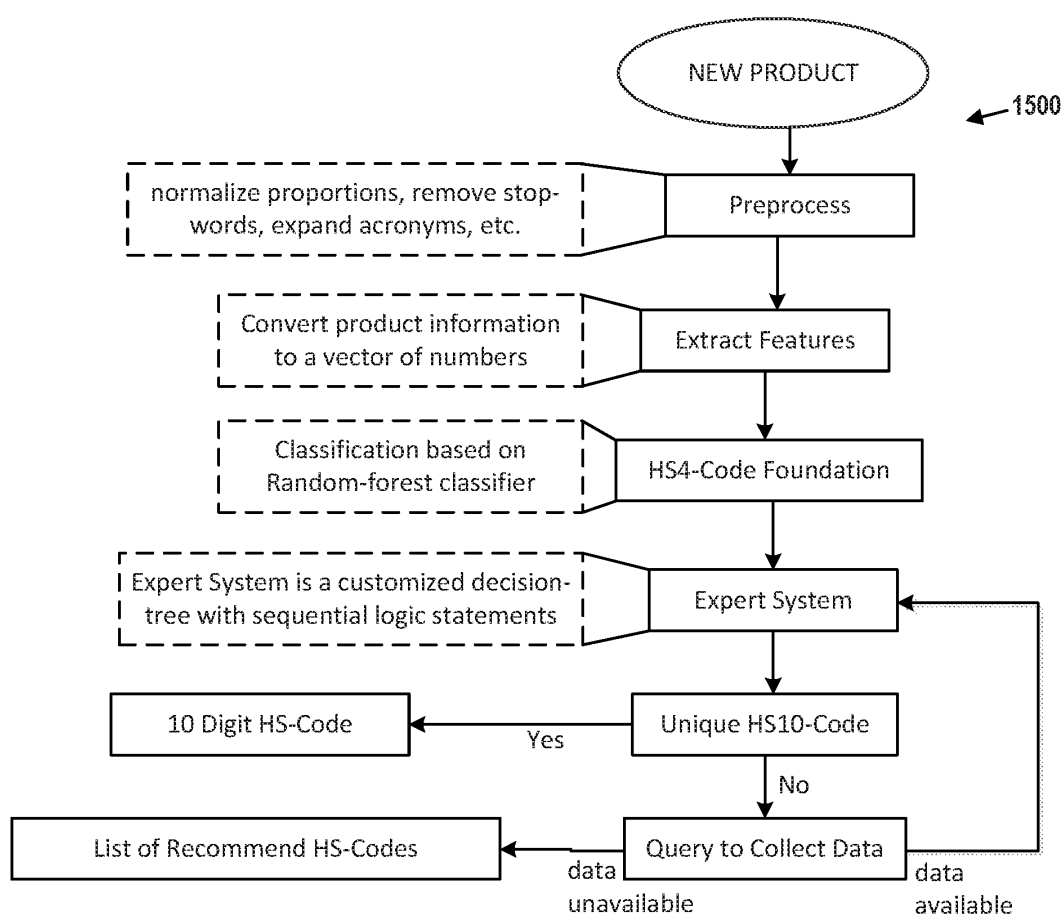


FIG. 15

19/24

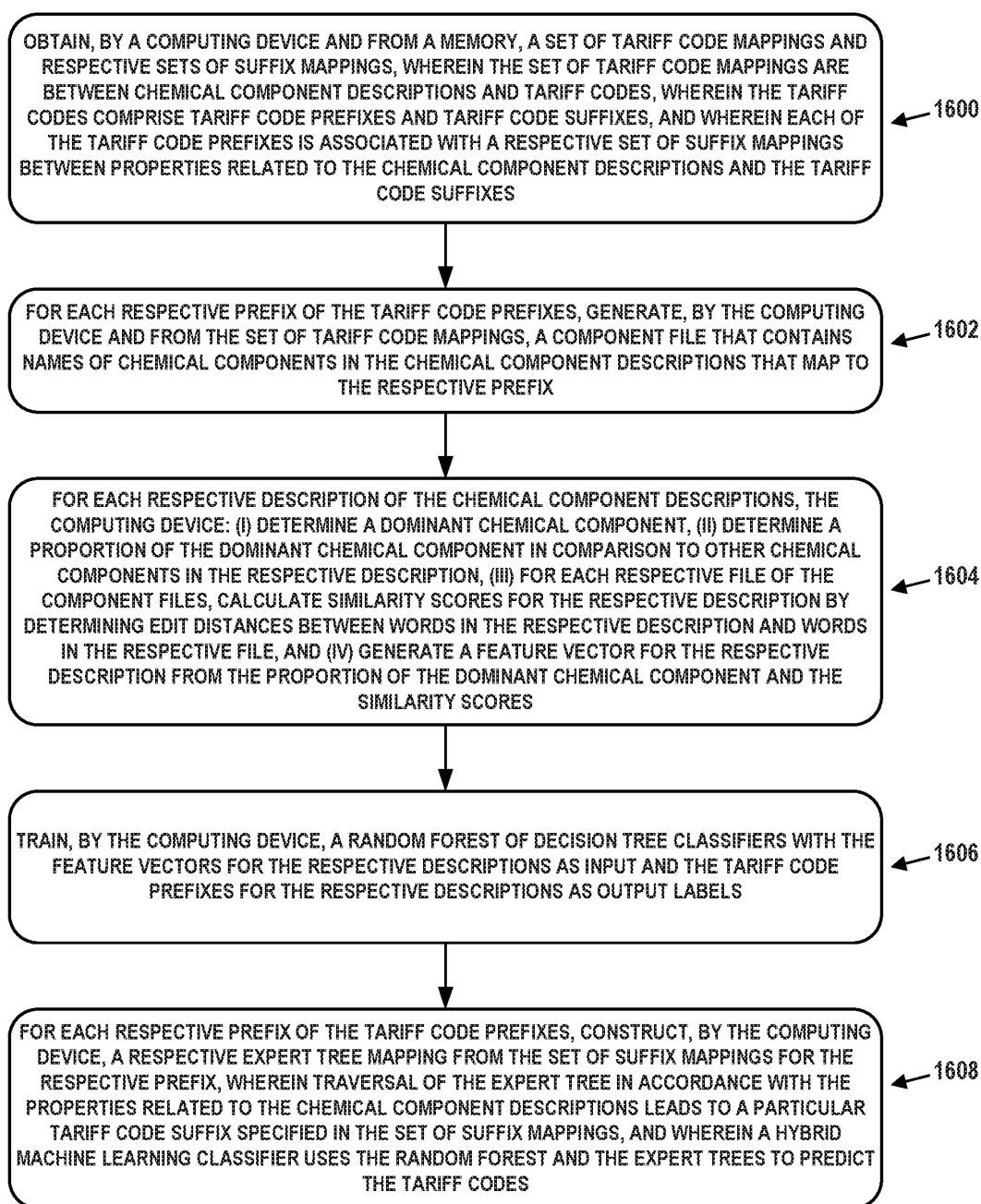


FIG. 16

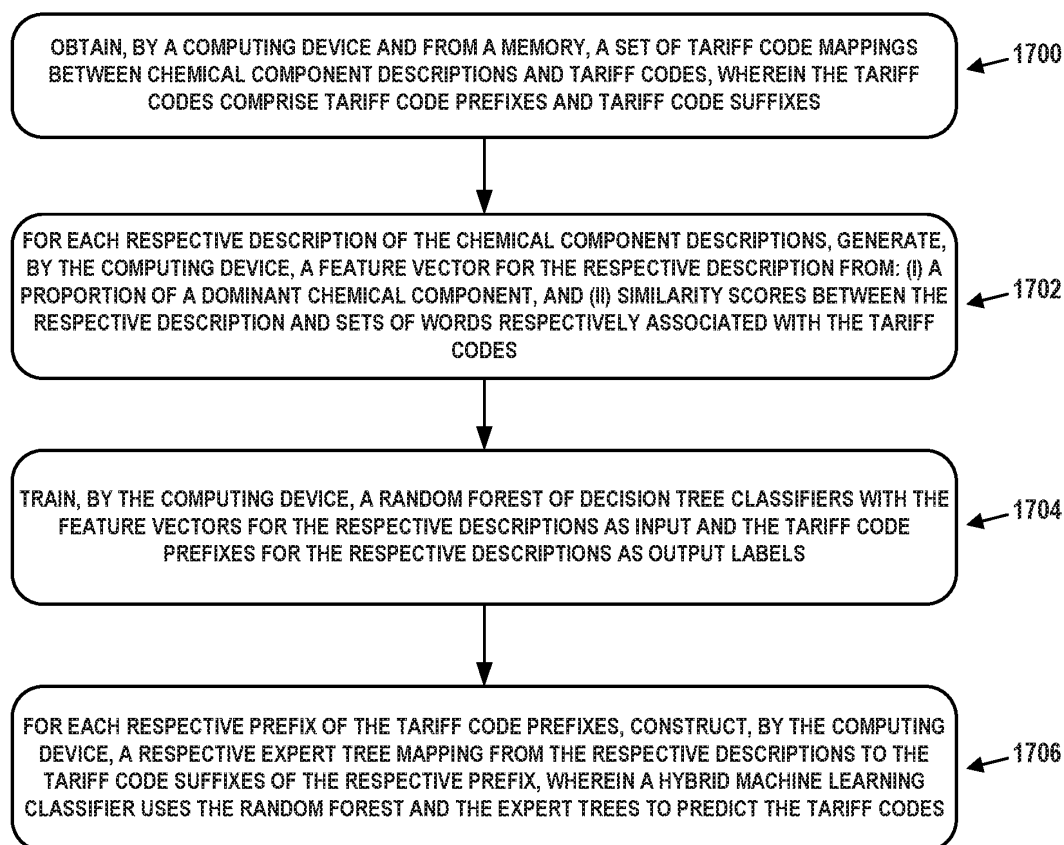


FIG. 17

21/24

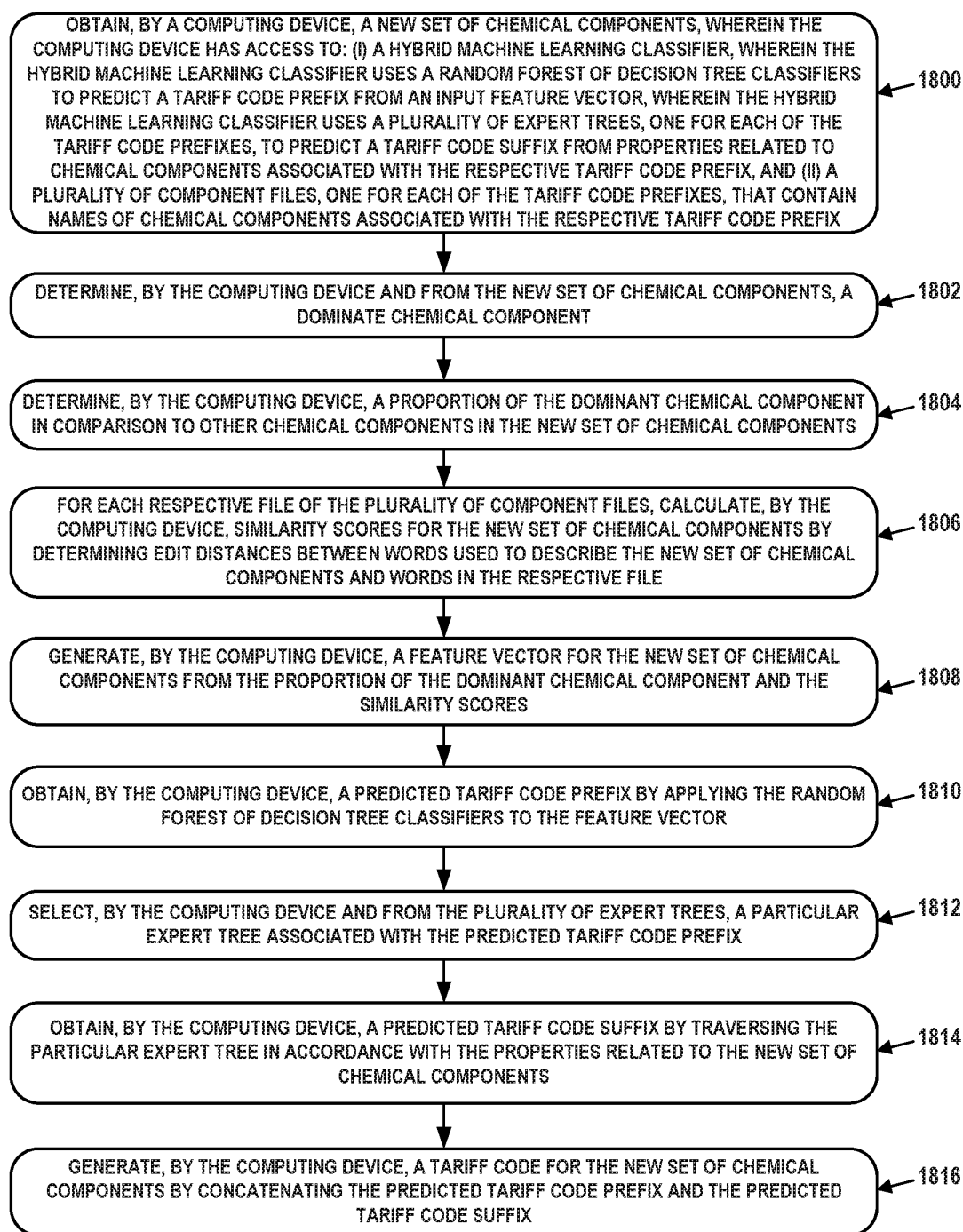


FIG. 18

22/24

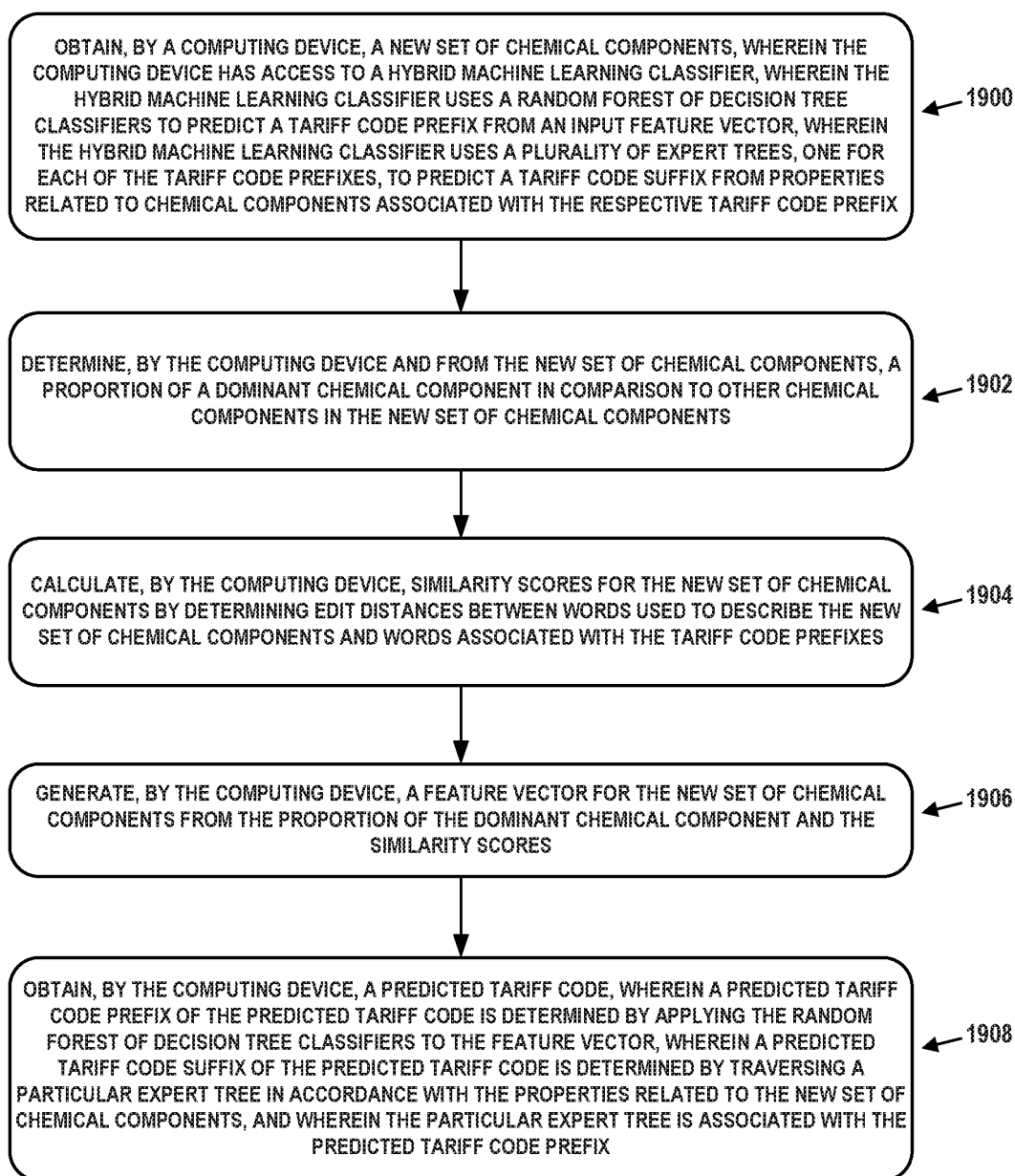


FIG. 19

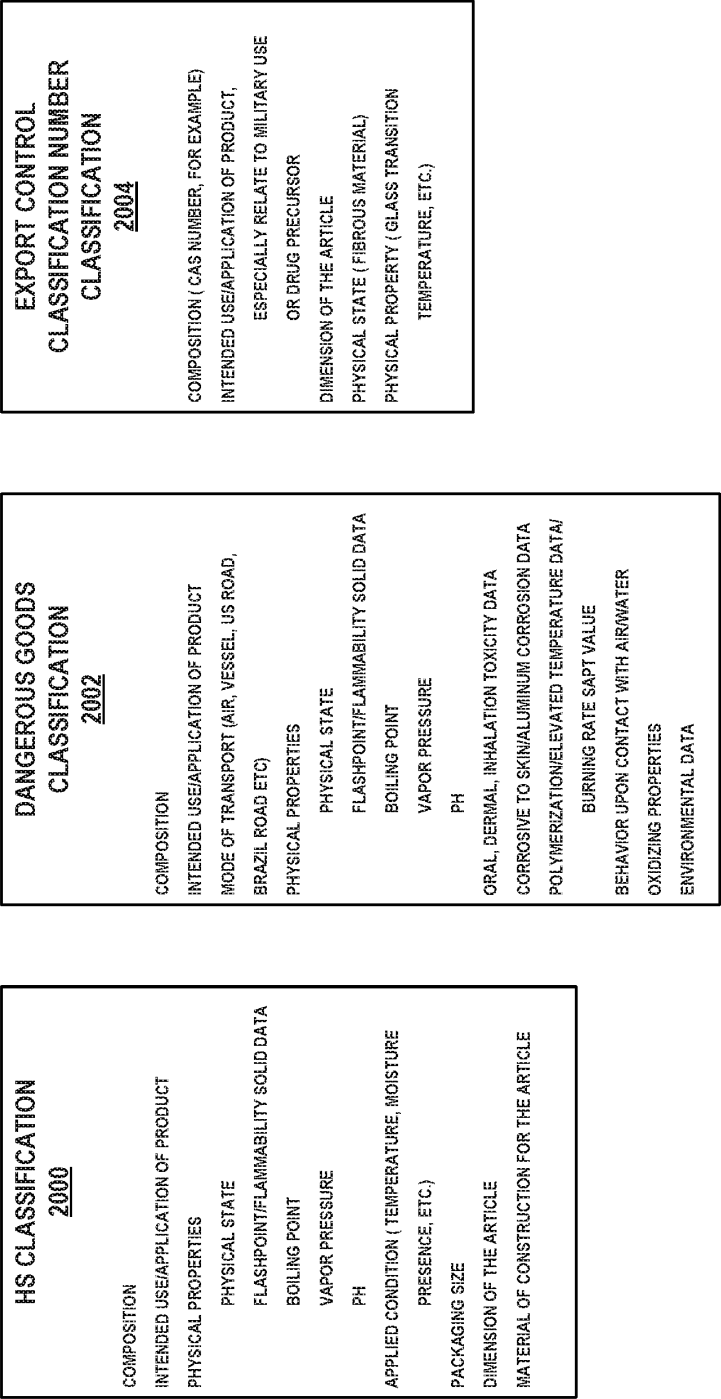


FIG. 20

24/24

**EXAMPLE DECISION TREE FOR EXPORT CONTROL CLASSIFICATION NUMBER
CLASSIFICATION 2100**

- a. Organic "fibrous or filamentary materials", having all of the following:
 - a.1. "Specific modulus" exceeding 12.7×10^6 m; and
 - a.2. "Specific tensile strength" exceeding 23.5×10^4 m;
 - b. Carbon "fibrous or filamentary materials", having all of the following:
 - b.1. "Specific modulus" exceeding 14.65×10^6 m; and
 - b.2. "Specific tensile strength" exceeding 26.82×10^4 m;
 - c. Inorganic "fibrous or filamentary materials", having all of the following:
 - c.1. "Specific modulus" exceeding 2.54×10^6 m; and
 - c.2. Melting, softening, decomposition or sublimation point exceeding 1,922 K (1,649 °C) in an inert environment;
 - d. "Fibrous or filamentary materials", having any of the following:
 - d.1. Composed of any of the following:
 - d.1.a. Polyetherimides controlled by 1C008.a; or
 - d.1.b. Materials controlled by 1C008.d to 1C008.f; or
 - d.2. Composed of materials controlled by 1C010.d.1.a or 1C010.d.1.b and "commingled" with other fibers controlled by 1C010.a, 1C010.b or 1C010.c;
 - e. Fully or partially resin-impregnated or pitch-impregnated "fibrous or filamentary materials" (prepregs), metal or carbon-coated "fibrous or filamentary materials" (preforms) or "carbon fiber preforms", having all of the following:
 - e.1. Having any of the following:
 - e.1.a. Inorganic "fibrous or filamentary materials" controlled by 1C010.c; or
 - e.1.b. Organic or carbon "fibrous or filamentary materials", having all of the following:
 - e.1.b.1. "Specific modulus" exceeding 10.15×10^6 m; and
 - e.1.b.2. "Specific tensile strength" exceeding 17.7×10^4 m; and
 - e.2. Having any of the following:
 - e.2.a. Resin or pitch, controlled by 1C008 or 1C009.b;
 - e.2.b. 'Dynamic Mechanical Analysis glass transition temperature (DMA T_g)' equal to or exceeding 453 K (180 °C) and having a phenolic resin; or
 - e.2.c. 'Dynamic Mechanical Analysis glass transition temperature (DMA T_g)' equal to or exceeding 505 K (232 °C) and having a resin or pitch, not specified by 1C008 or 1C009.b, and not being a phenolic resin;
- Non-fluorinated polymeric substances
- a. Imides as follows:
 - a.1. Bismaleimides;
 - a.2. Aromatic polyamide-imides (PAI) having a 'glass transition temperature (T_g)' exceeding 563 K (290 °C);
 - a.3. Aromatic polyimides having a 'glass transition temperature (T_g)' exceeding 505 K (232 °C);
 - a.4. Aromatic polyetherimides having a 'glass transition temperature (T_g)' exceeding 563 K (290 °C);
 - d. Polyarylene ketones;
 - e. Polyarylene sulphides, where the arylene group is biphenylene, triphenylene or combinations thereof;
 - f. Polybiphenylenethersulphone having a 'glass transition temperature (T_g)' exceeding 563 K (290 °C).

FIG. 21