
Text Collection
Clustering & Summarization
clusters, summaries
Pre-Processing
Display

Figure 1

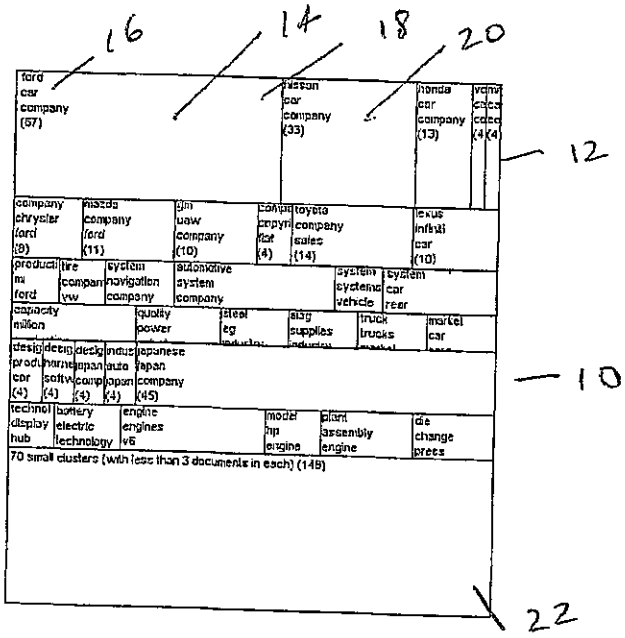

16
14
18
20
ford car company (67)
nissan car company (33)
honda car company (13)
company chrysler ford (9)
nissan company ford (11)
gm uaw company (10)
toyota company sales (14)
lexus infiniti car (10)
tire
system navigation company
automotive system company
system
car
capacity million
quality power
steel
supplies
truck trucks
market car
desig
deep
desig
japan
indus
auto
japanese japan company (45)
technol display hub
battery electric technology
engine engines v6
model hp engine
plant assembly engine
die change press
70 small clusters (with less than 3 documents in each) (149)
12
10
22

Figure 2

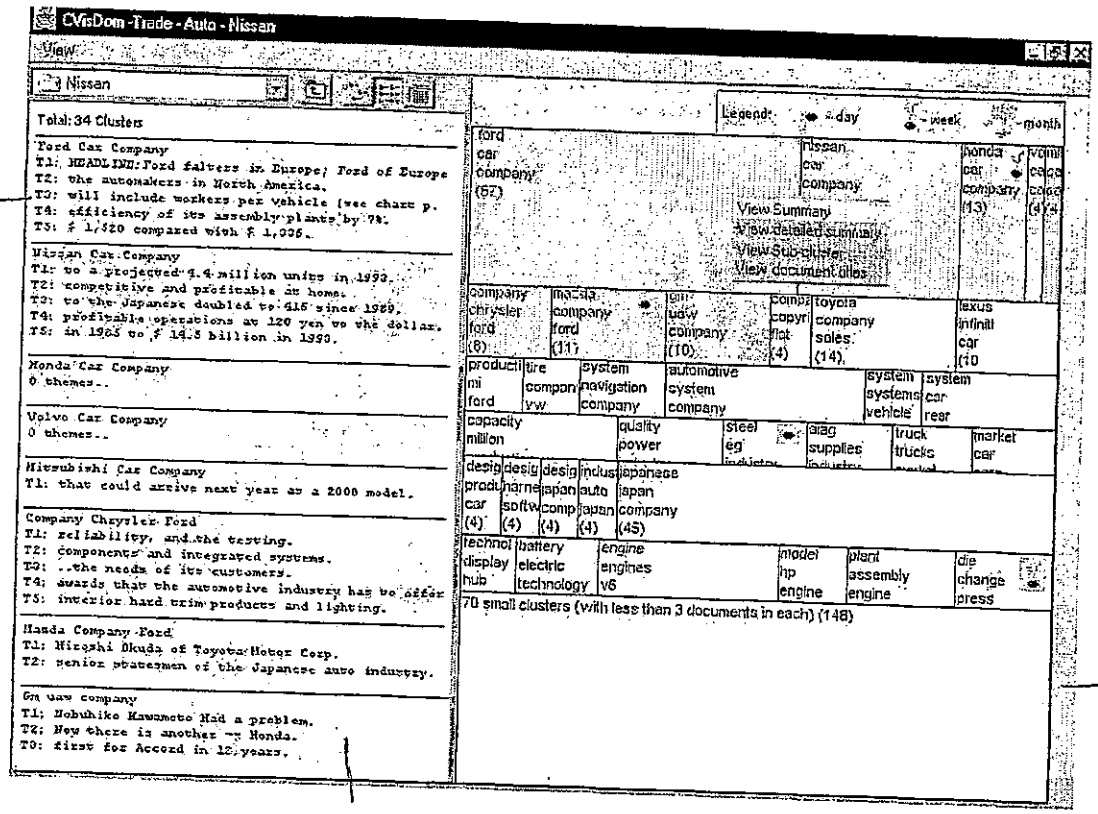

CVisDom - Trade - Auto - Nissan
View
Nissan
Total: 34 Clusters
Ford Car Company
T1: HEADLINE: Ford falters in Europe; Ford of Europe
T2: the automakers in North America.
T3: will include workers per vehicle (see chart p.
T4: efficiency of its assembly plants by 7%.
T5: $ 1,920 compared with $ 1,935.
Nissan Car Company
T1: to a projected 4.4 million units in 1993.
T2: competitive and profitable at home.
T3: to the Japanese doubled to 415 since 1989.
T4: profitable operations at 120 yen to the dollar.
T5: in 1965 to $ 14.5 billion in 1993.
Honda Car Company
0 themes..
Volvo Car Company
0 themes..
Mitsubishi Car Company
T1: that could arrive next year as a 2000 model.
Company Chrysler Ford
T1: reliability, and the testing.
T2: components and integrated systems.
T3: ..the needs of its customers.
T4: awards that the automotive industry has to offer
T5: interior hard trim products and lighting.
Honda Company Ford
T1: Hiroshi Okuda of Toyota Motor Corp.
T2: senior statesmen of the Japanese auto industry.
Gm uaw company
T1: Nobuhiko Kawamoto Had a problem.
T2: Now there is another -- Honda.
T3: first for Accord in 12 years.
Legend: day week month
ford car company (67)
nissan car company
honda car company (13)
View Summary
View detailed summary
View Sub-cluster
View document titles
company chrysler ford (8)
mazda company ford (11)
gm uaw company (10)
compa copyri fiat (4)
toyota company sales (14)
lexus infiniti car (10
produci mi ford
tire compan vw
system navigation company
automotive system company
system systems vehicle
system car rear
capacity million
quality power
steel eg
drag supplies
truck trucks
market car
desig produ car (4)
desig harne softw (4)
desig japan comp (4)
indus auto japan (4)
japanese japan company (45)
technol display hub
battery electric technology
engine engines v6
model hp engine
plant assembly engine
die change press
70 small clusters (with less than 3 documents in each) (148)
26
10
24

Figure 3

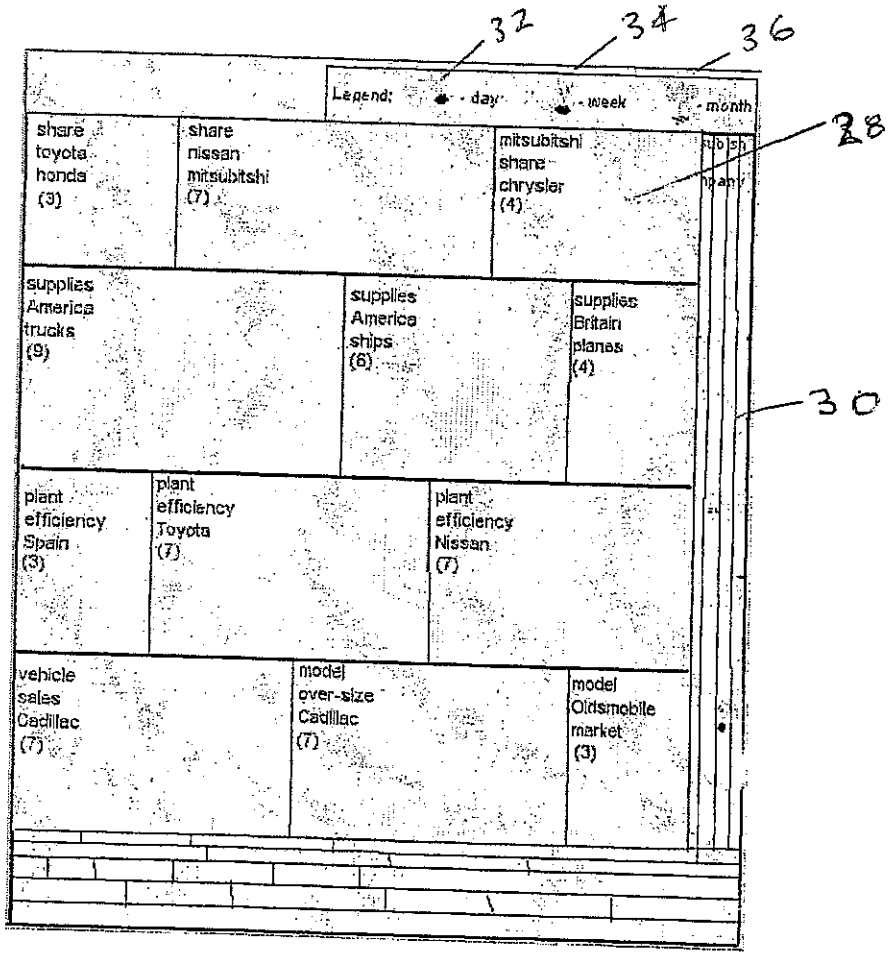

32
34
36
Legend:
- day
- week
- month
share
toyota
honda
(3)
share
nissan
mitsubitshi
(7)
mitsubitshi
share
chrysler
(4)
38
supplies
America
trucks
(9)
supplies
America
ships
(6)
supplies
Britain
planes
(4)
30
plant
efficiency
Spain
(3)
plant
efficiency
Toyota
(7)
plant
efficiency
Nissan
(7)
vehicle
sales
Cadillac
(7)
model
over-size
Cadillac
(7)
model
Oldsmobile
market
(3)

Figure 4

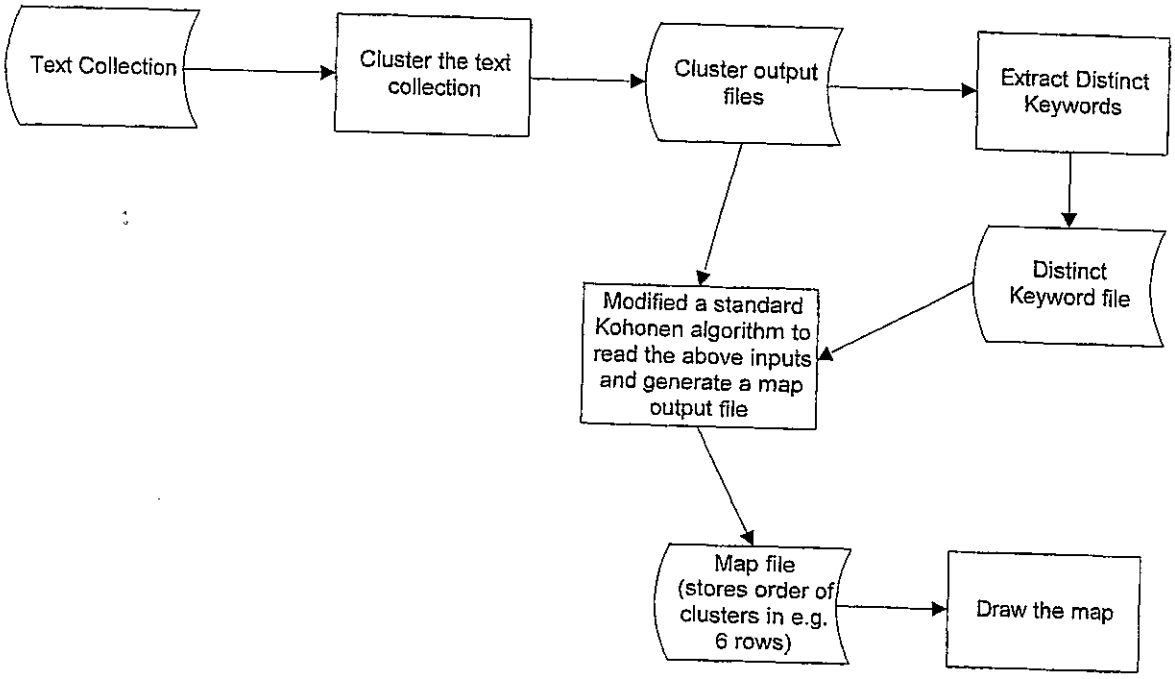
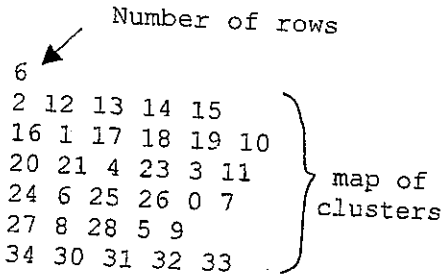

Text Collection
Cluster the text collection
Cluster output files
Extract Distinct Keywords
Distinct Keyword file
Modified a standard Kohonen algorithm to read the above inputs and generate a map output file
Map file (stores order of clusters in e.g. 6 rows)
Draw the map

Figure 5


Number of rows
6
2 12 13 14 15
16 1 17 18 19 10
20 21 4 23 3 11
24 6 25 26 0 7
27 8 28 5 9
34 30 31 32 33
map of clusters

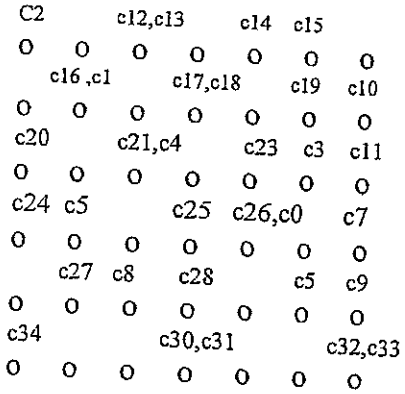

C2
c12,c13
c14
c15
c16,c1
c17,c18
c19
c10
c20
c21,c4
c23
c3
c11
c24
c5
c25
c26,c0
c7
c27
c8
c28
c5
c9
c34
c30,c31
c32,c33

Figure 6

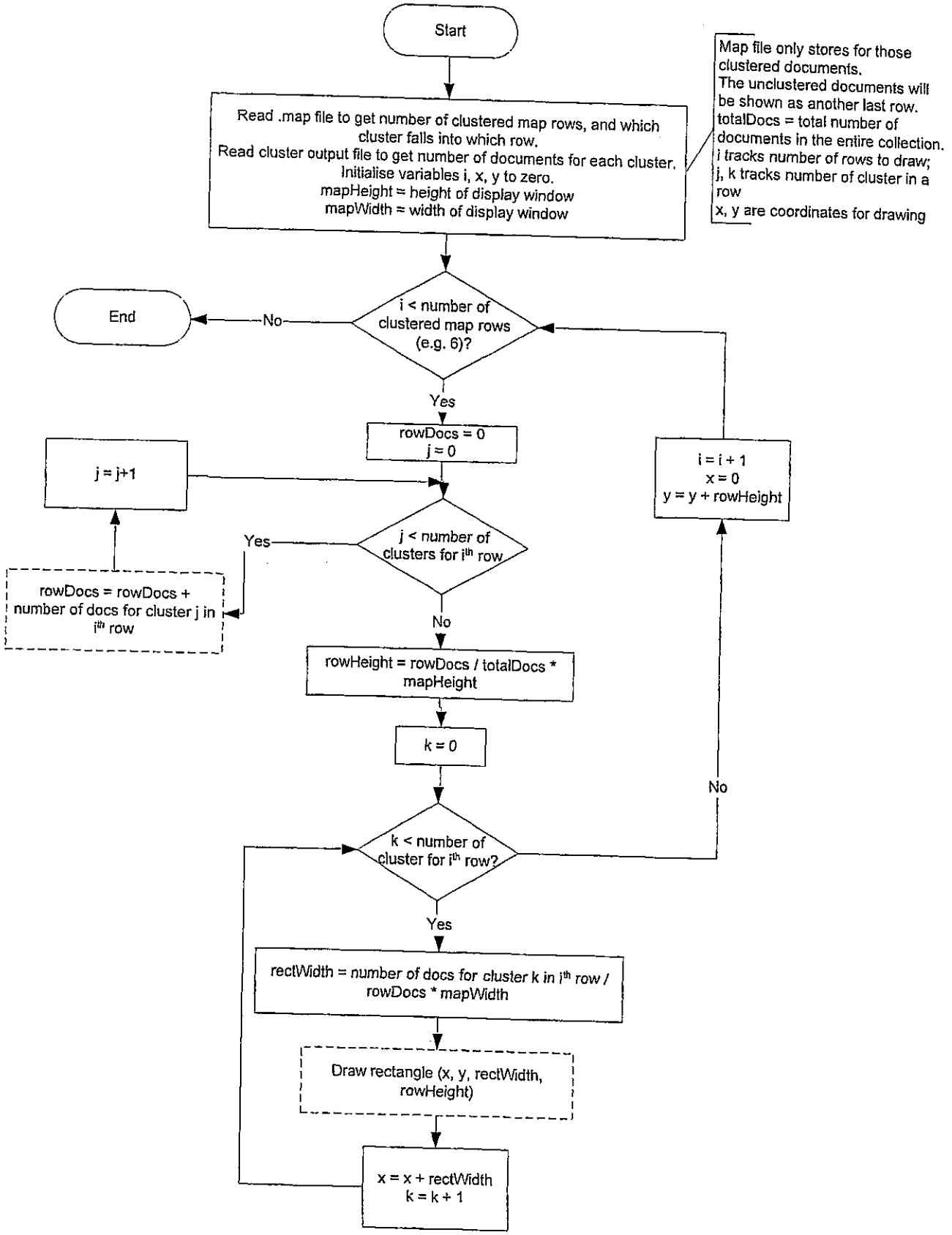

Start
Read .map file to get number of clustered map rows, and which cluster falls into which row.
Read cluster output file to get number of documents for each cluster.
Initialise variables i, x, y to zero.
mapHeight = height of display window
mapWidth = width of display window
Map file only stores for those clustered documents.
The unclustered documents will be shown as another last row.
totalDocs = total number of documents in the entire collection.
i tracks number of rows to draw;
j, k tracks number of cluster in a row
x, y are coordinates for drawing
i < number of clustered map rows (e.g. 6)?
No
End
Yes
rowDocs = 0
j = 0
j < number of clusters for ith row
Yes
rowDocs = rowDocs + number of docs for cluster j in ith row
j = j+1
No
rowHeight = rowDocs / totalDocs * mapHeight
k = 0
k < number of cluster for ith row?
No
Yes
rectWidth = number of docs for cluster k in ith row / rowDocs * mapWidth
Draw rectangle (x, y, rectWidth, rowHeight)
x = x + rectWidth
k = k + 1
i = i + 1
x = 0
y = y + rowHeight

Figure 7

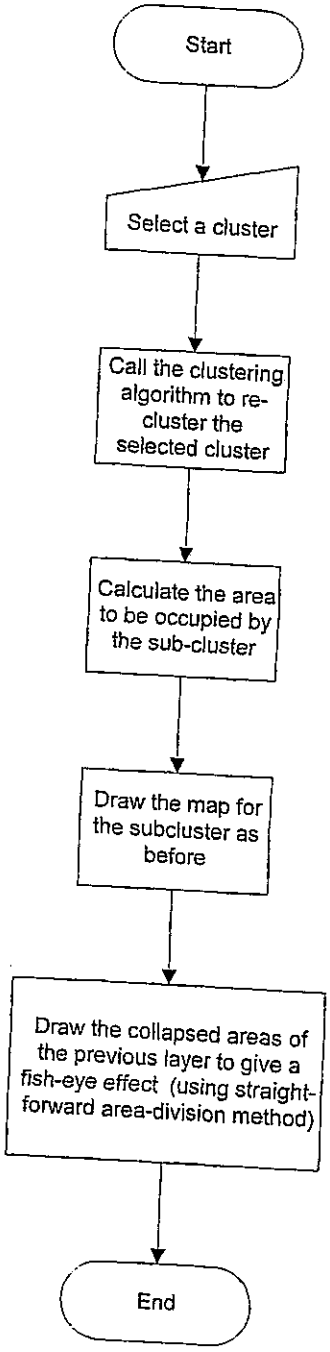

Start
Select a cluster
Call the clustering algorithm to re-cluster the selected cluster
Calculate the area to be occupied by the sub-cluster
Draw the map for the subcluster as before
Draw the collapsed areas of the previous layer to give a fish-eye effect (using straight-forward area-division method)
End

Figure 8

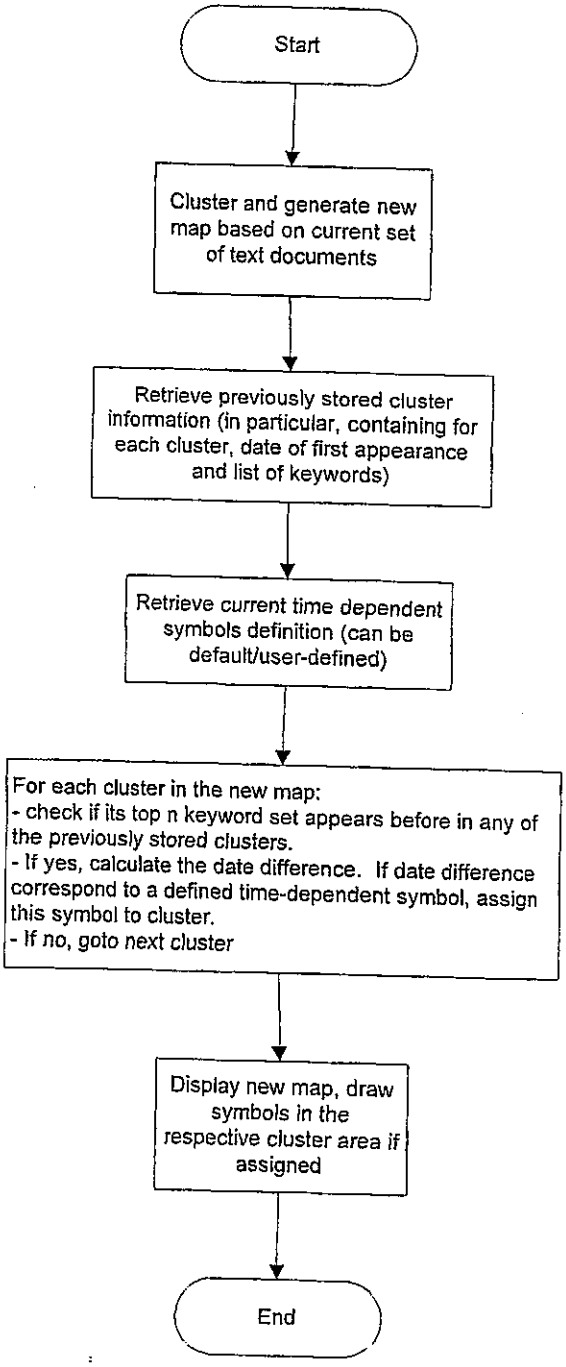

Start
Cluster and generate new map based on current set of text documents
Retrieve previously stored cluster information (in particular, containing for each cluster, date of first appearance and list of keywords)
Retrieve current time dependent symbols definition (can be default/user-defined)
For each cluster in the new map:
- check if its top n keyword set appears before in any of the previously stored clusters.
- If yes, calculate the date difference. If date difference correspond to a defined time-dependent symbol, assign this symbol to cluster.
- If no, goto next cluster
Display new map, draw symbols in the respective cluster area if assigned
End

Figure 9

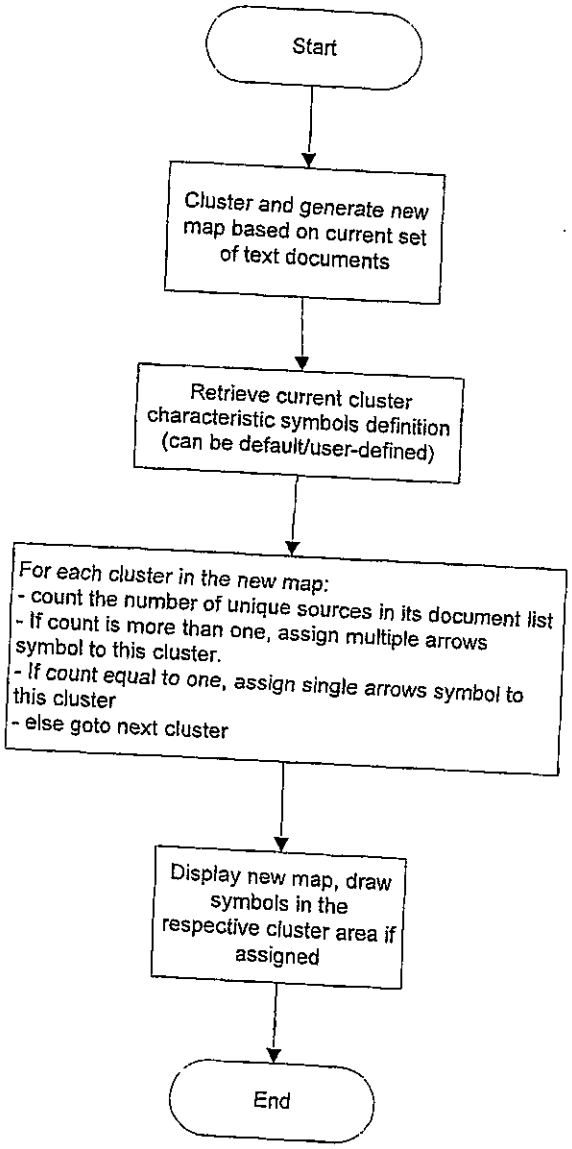

Start
Cluster and generate new map based on current set of text documents
Retrieve current cluster characteristic symbols definition (can be default/user-defined)
For each cluster in the new map:
- count the number of unique sources in its document list
- If count is more than one, assign multiple arrows symbol to this cluster.
- If count equal to one, assign single arrows symbol to this cluster
- else goto next cluster
Display new map, draw symbols in the respective cluster area if assigned
End

Figure 10

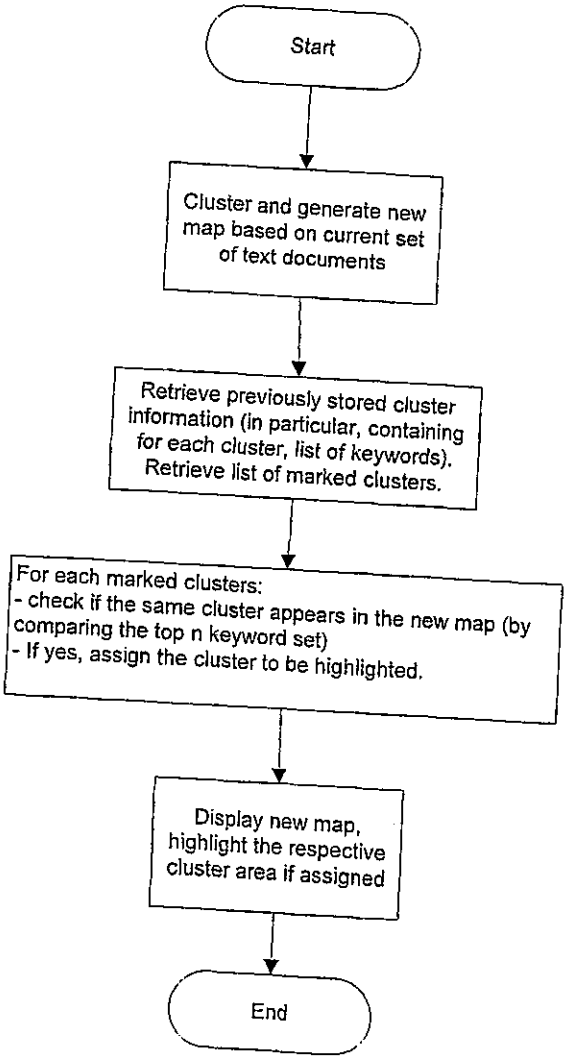

Start
Cluster and generate new map based on current set of text documents
Retrieve previously stored cluster information (in particular, containing for each cluster, list of keywords). Retrieve list of marked clusters.
For each marked clusters:
- check if the same cluster appears in the new map (by comparing the top n keyword set)
- If yes, assign the cluster to be highlighted.
Display new map, highlight the respective cluster area if assigned
End

Figure 11

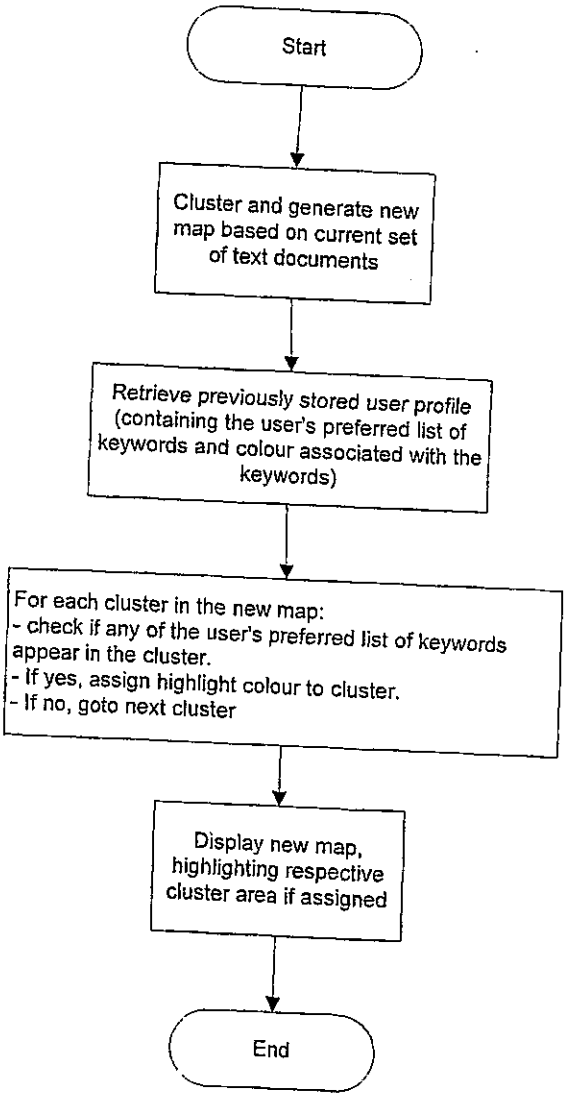

Start
Cluster and generate new map based on current set of text documents
Retrieve previously stored user profile (containing the user's preferred list of keywords and colour associated with the keywords)
For each cluster in the new map:
- check if any of the user's preferred list of keywords appear in the cluster.
- If yes, assign highlight colour to cluster.
- If no, goto next cluster
Display new map, highlighting respective cluster area if assigned
End

Figure 12