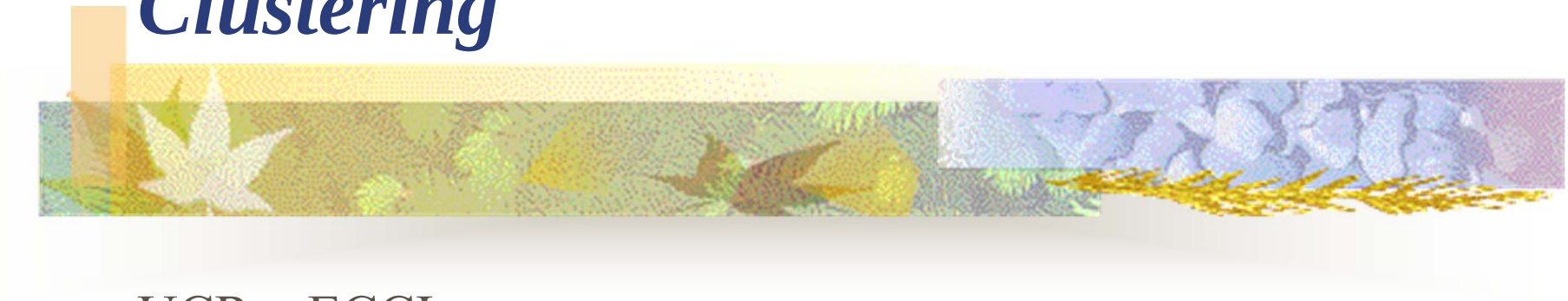
# Interfaces de Visualización de *Clustering*



UCR – ECCI

CI-2414 Recuperación de Información

Prof. Kryscia Daviana Ramírez Benavides

# Introducción

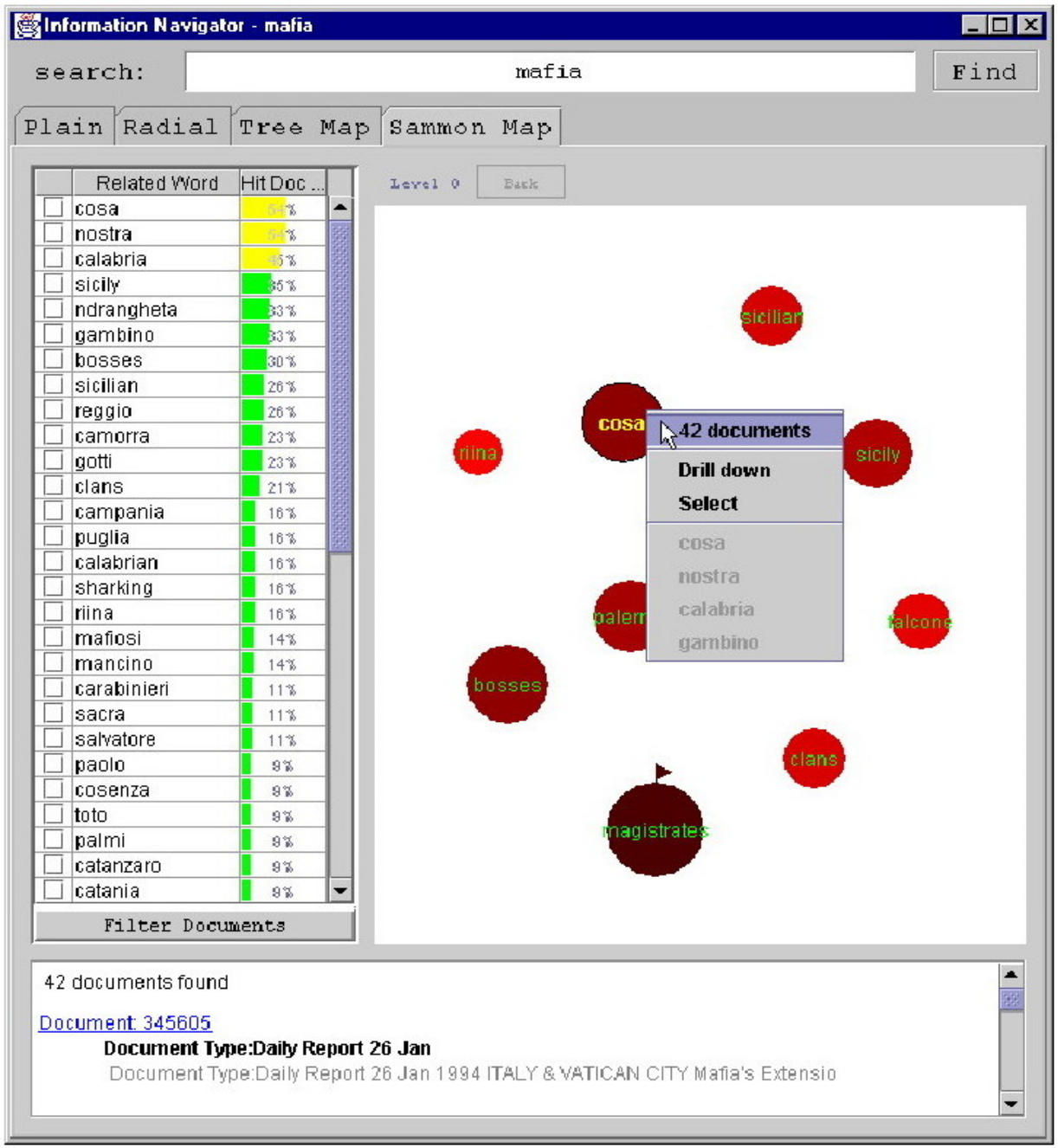
- Los mecanismos convencionales de una búsqueda tienen baja precisión.
- Un problema común con esto es que los usuarios deben navegar a través de muchos documentos irrelevantes antes de encontrar el tipo de documento que le interesa.
- Aún con algoritmos avanzados de *ranking*, no se puede saber de antemano que tipo de documentos prefiere el usuario.
- *Interfases Gráficas basadas en Clustering* pueden ayudar al usuario a encontrar más fácilmente lo que busca.

# Nuevos Paradigmas en Visualización de Información

- *Sammon Map*.
- *Tree-Map Visualization*.
- *Radial Interactive Visualization*.

## *Sammon Map*

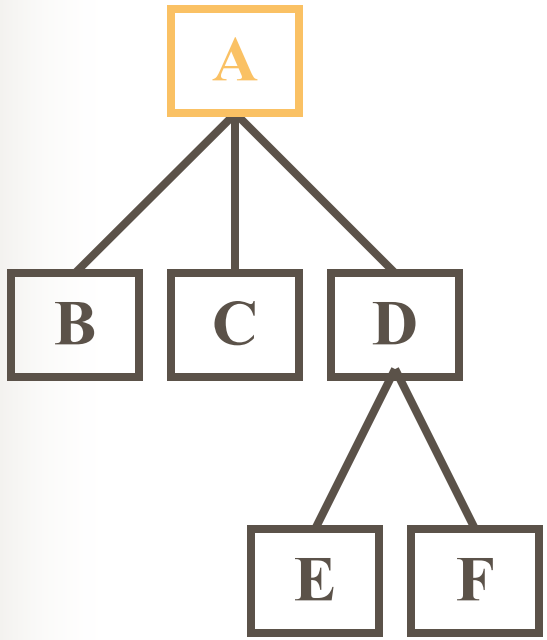
- *Sammon map* genera una localización en dos dimensiones de los clusters.
- Este mapa se calcula usando una búsqueda de gradiante iterativo.
- El objetivo del algoritmo es representar puntos en un espacio *n* dimensiones, normalmente 2 dimensiones, mientras intenta conservar las distancia pares entre los objetos.
- Utiliza el algoritmo de *clustering* Buckshot.
- Despliega documentos en *clustering* por *keywords*.
- Al dar click a cualquier *cluster* muestra el número de documentos que contiene, y el *ranking* de las *keywords* usadas para formar el *cluster*.
- Al dar click en el número de documentos muestra las *keywords* y sus porcentajes, y un resumen de los documentos.


Information Navigator - mafia
search:
mafia
Find
Plain
Radial
Tree Map
Sammon Map
Related Word
Hit Doc ...
cosa
nostra
calabria
sicily
ndrangheta
gambino
bosses
sicilian
reggio
camorra
gotti
clans
campania
puglia
calabrian
sharking
riina
mafiosi
mancino
carabinieri
sacra
salvatore
paolo
cosenza
toto
palmi
catanzaro
catania
Filter Documents
Level 0
Back
sicilian
cosa
riina
sicily
palerm
falcone
bosses
clans
magistrates
42 documents
Drill down
Select
cosa
nostra
calabria
gambino
42 documents found
Document: 345605
Document Type:Daily Report 26 Jan
Document Type:Daily Report 26 Jan 1994 ITALY & VATICAN CITY Mafia's Extensio

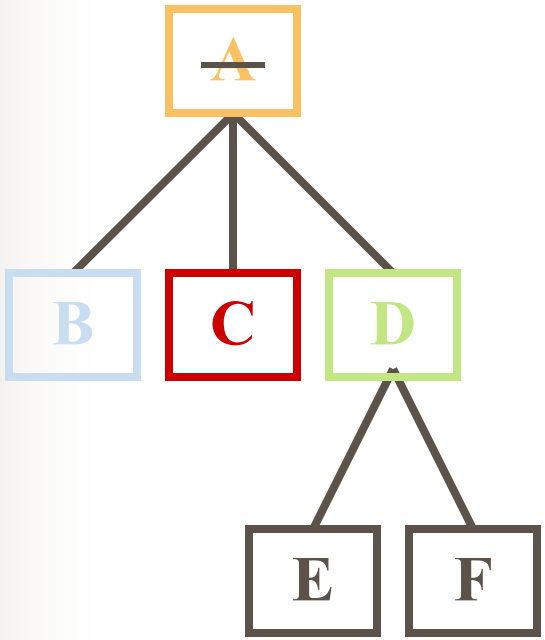
## *Tree-Map Visualization*

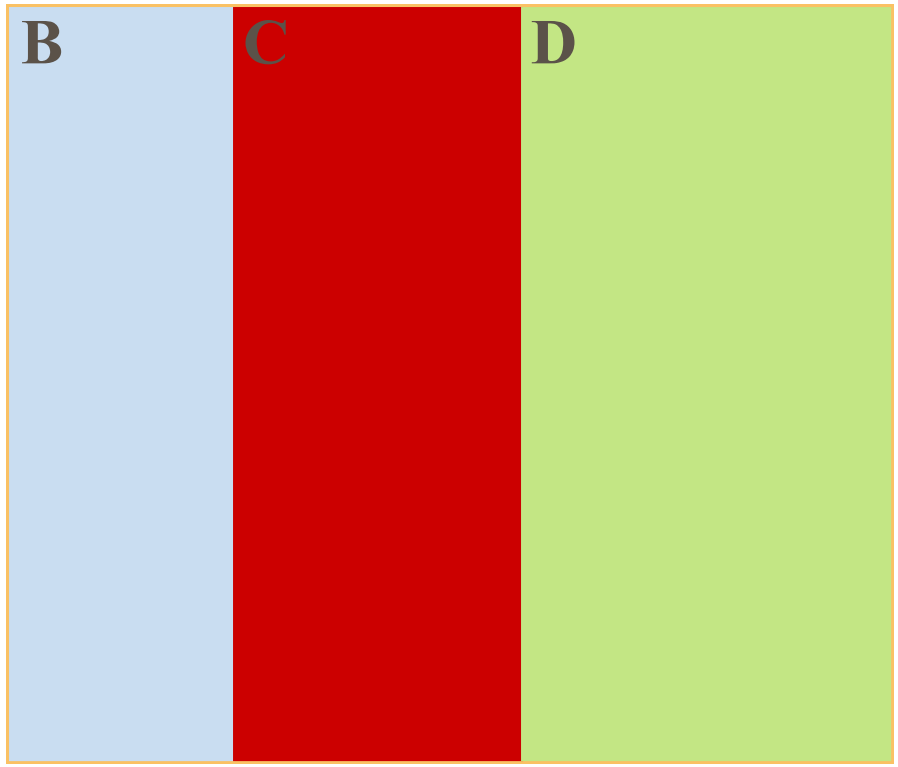
- Representación jerárquica de los *clusters*.
- Los *clusters* se visualizan mediante rectángulos.
- *Clusters* similares se agrupan en Súper *Clusters*.
- Utiliza algoritmos de *clustering* jerárquicos. Aunque puede usar no jerárquicos.

# *Tree-Map Visualization* (cont.)

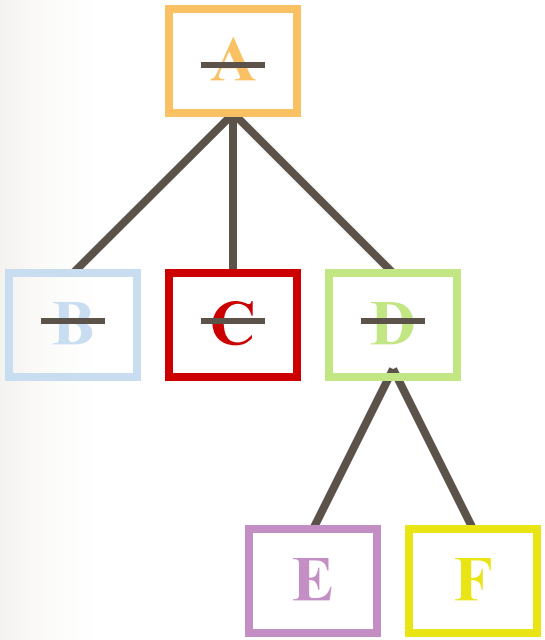


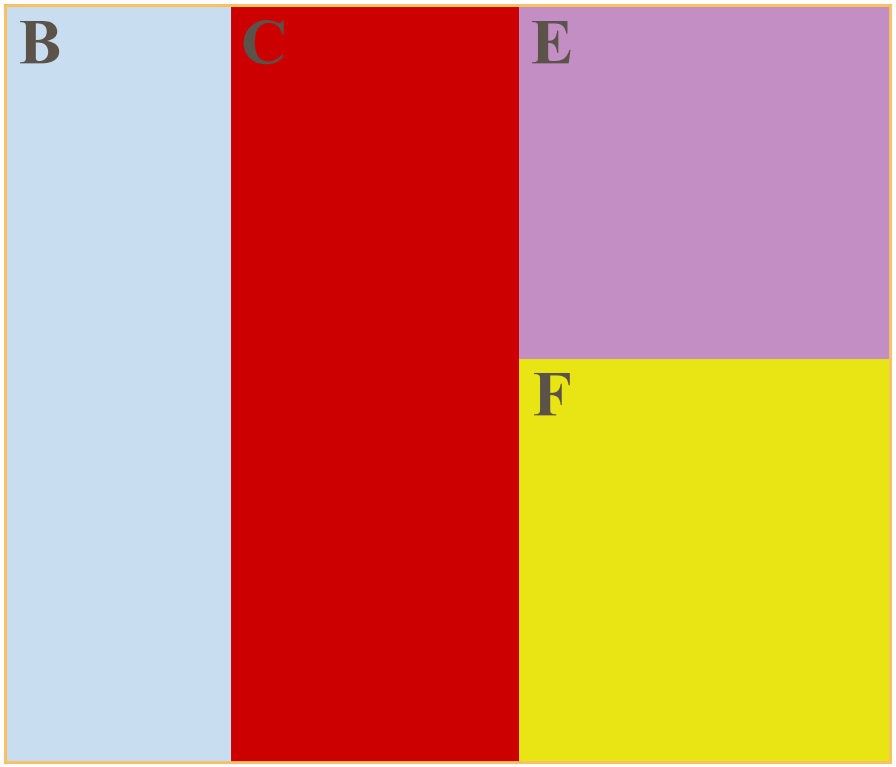

# *Tree-Map Visualization* (cont.)

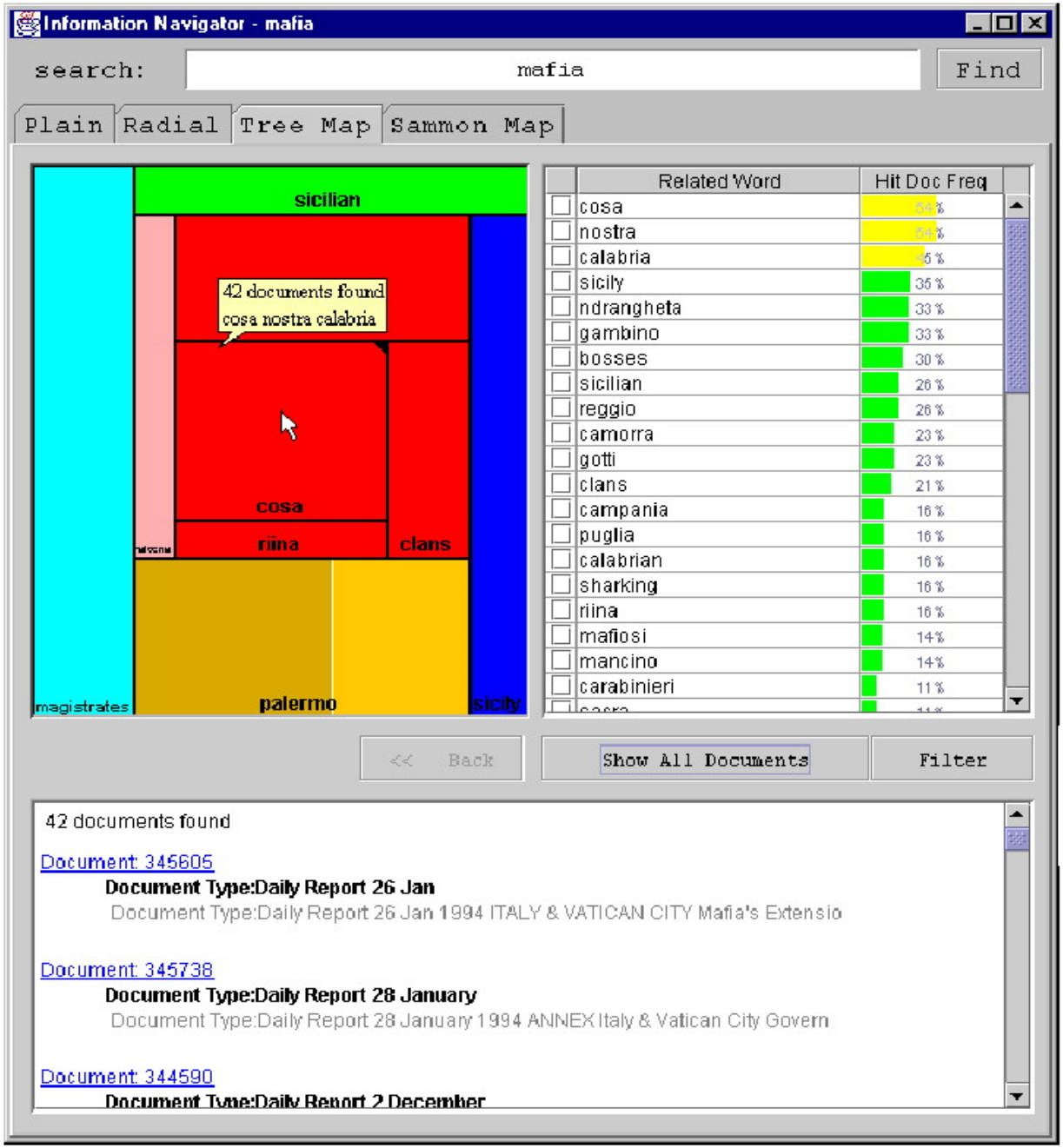


# *Tree-Map Visualization* (cont.)



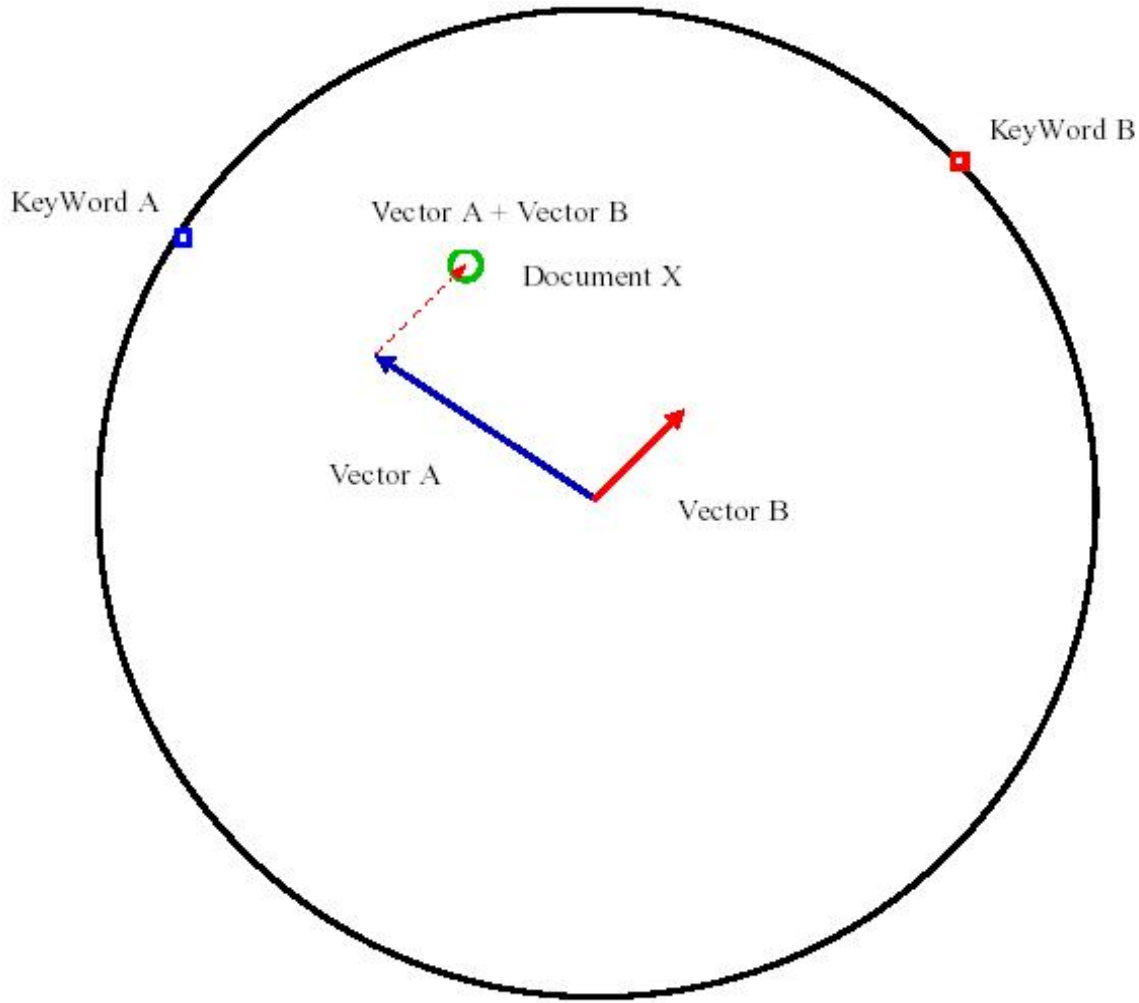



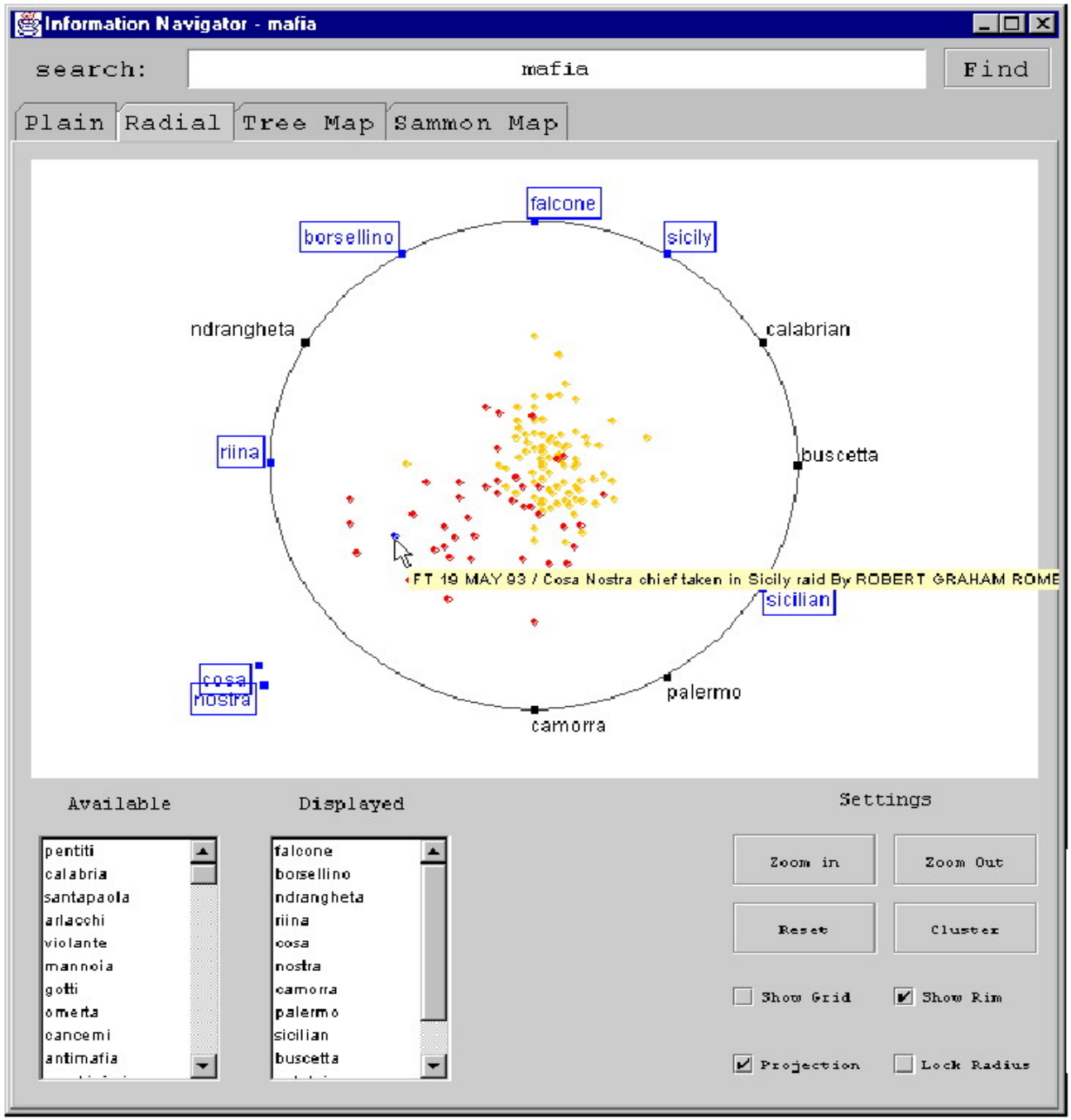

Information Navigator - mafia
search:
mafia
Find
Plain
Radial
Tree Map
Sammon Map
sicilian
42 documents found
cosa nostra calabria
cosa
riina
clans
magistrates
palermo
sicily
Related Word
Hit Doc Freq
cosa
nostra
calabria
sicily
ndrangheta
gambino
bosses
sicilian
reggio
camorra
gotti
clans
campania
puglia
calabrian
sharking
riina
mafiosi
mancino
carabinieri
<< Back
Show All Documents
Filter
42 documents found
Document: 345605
Document Type:Daily Report 26 Jan
Document Type:Daily Report 26 Jan 1994 ITALY & VATICAN CITY Mafia's Extensio
Document: 345738
Document Type:Daily Report 28 January
Document Type:Daily Report 28 January 1994 ANNEX Italy & Vatican City Govern
Document: 344590
Document Type:Daily Report 2 December

## *Radial Interactive Visualization*

- Nodos con *keywords* son colocados alrededor de un círculo.
- Documentos están representados por puntos en el interior del círculo.
- Entre más relacionado un documento con un *keyword* más cercano estará de este.
- Utiliza el algoritmo de *clustering* Buckshot.
- Al dar click en una palabra resaltarán los documentos que contienen esa palabra. Al dar click en un punto mostrarán las palabras que contiene.
- Algoritmo:
  - Sea P1 la posición del *keyword* A en el círculo.
  - Sea P2 la posición del *keyword* B en el círculo.
  - Se utiliza la matriz de pesos de *keywords* para cada documento:

    (P1 * wij,P2 * wij).


KeyWord A
KeyWord B
Vector A + Vector B
Document X
Vector A
Vector B


Information Navigator - mafia
search:
mafia
Find
Plain
Radial
Tree Map
Sammon Map
falcone
borsellino
sicily
ndrangheta
calabrian
riina
buscetta
FT 19 MAY 93 / Cosa Nostra chief taken in Sicily raid By ROBERT GRAHAM ROME
sicilian
cosa
nostra
palermo
camorra
Available
Displayed
Settings
pentiti
calabria
santapaola
arlacchi
violante
mannoia
gotti
omerta
cancemi
antimafia
falcone
borsellino
ndrangheta
riina
cosa
nostra
camorra
palermo
sicilian
buscetta
Zoom in
Zoom Out
Reset
Cluster
Show Grid
Show Rim
Projection
Lock Radius

# Características Comparativas

- Visualizaciones basadas en *clusters* dan un panorama más amplio del conjunto de resultados.
- *Sammon Map* se enfoca más en la relación que existe en un *cluster* y otro.
- *Tree Map* es más explícito en cuanto al tamaño de los *cluster* y su estructura jerárquica.
- *Radial Visualization* permite enfocarse en formar subconjuntos de *keywords*.

# Facilidades de cada Interfaz

- *Sammon Map* guía en el análisis, permite reagrupar subconjuntos, y gradualmente acercar al tipo de documentos de interés.
- *Tree Map* permite enfocarse en *keywords* de ciertos documentos de interés, para formular una búsqueda más productiva.
- *Radial Visualization* es apropiado si el usuario esta familiarizado con los *keywords* del área de su interés.

# Conclusiones

- Contribuye a la visualización y navegación de un conjunto de documentos retornados mediante:
  - Identificación de *keywords* relevantes en un conjunto de documentos.
  - Desechar rápidamente *clusters* irrelevantes.
  - Operaciones de "*Drill Down*" en *clusters* relevantes.
  - Construcción de grupos personalizados.

# Referencias Bibliográficas

- La información fue tomada de:
  - Libro de texto del curso.
  - Presentaciones del curso RI del estudiante Randall Mora Jiménez. Universidad de Costa Rica, 2003.
  - A Visualization Interface for Document Searching and Browsing. Matthew Carey Frank Kriwaczek Stefan M Ruger Multimedia Knowledge http://km.doc.ic.ac.uk.