

Sistemas de Recuperación de Información

Aplicaciones del
agrupamiento y
de la clasificación
en la recuperación
de información
en la Web

Autores

Laura Victoria Riera Pérez

Marcos Manuel Tirador del Riego

Índice general

1. Agrupamiento	1
1.1. Algunas definiciones	1
1.2. Flat clustering	1
1.3. Hierarchical clustering	1
1.4. Aplicaciones a la RI	2
1.5. Ventajas	2
1.6. Desventajas	2
2. Clasificación	2
2.1. Aplicaciones a la RI	2
2.2. Ventajas	2
2.3. Desventajas	2
3. Agrupamiento vs. Clasificación	2
4. Ejemplos de aplicación	2

1. Agrupamiento

- Aprendizaje no supervisado
- Problema que resuelve

1.1. Algunas definiciones

- Flat clustering
- Hierarchical clustering
- Hard clustering
- Soft clustering
- Hipótesis de agrupamiento
- Cardinalidad

1.2. Flat clustering

- Medida de similitud:
- Medidas de evaluación:
 - Criterio interno de calidad
 - Criterio interno de calidad
 - Pureza
 - Índice de frontera?
 - Medida F
- Algoritmos:
 - K-means
 - EM (generalización de K-means)

1.3. Hierarchical clustering

- Hierarchical agglomerative clustering
- Medidas de similitud:
 - Single link clustering
 - Complete link clustering
 - Centroid clustering
- Evaluación de calidad:
 - Group average link
 - Método de Ward
- Divisive clustering
- Cluster labeling
- Algoritmos:
 - Algoritmo HAC
 - Divisive Clustering

1.4. Aplicaciones a la RI

- Search result clustering
- Scatter-Gather
- Collection clustering
- Language modeling
- Cluster-based retrieval

1.5. Ventajas

1.6. Desventajas

2. Clasificación

- Aprendizaje supervisado
- Problema que resuelve
- Rule-based classification
- Statistical classification
- Feature selection
- Medidas de evaluación:
 - Fitting
 - Precisión
 - Recuperado
 - Medida F (balanceada)
 - Classification accuracy
- Algoritmos:
 - Naive Bayes
 - K-Nearest Neighbours

2.1. Aplicaciones a la RI

- Standing queries
- Spam filtering

2.2. Ventajas

2.3. Desventajas

3. Agrupamiento vs. Clasificación

4. Ejemplos de aplicación