



Universidad de Granada

[decsai.ugr.es](http://decsai.ugr.es)

# Modelos de ciencia de datos no numéricos: Aplicaciones en redes sociales, web y gestión de procesos



**Departamento de Ciencias de la  
Computación e Inteligencia Artificial**



Universidad de Granada

[decsai.ugr.es](http://decsai.ugr.es)

## Bloque II: Minería de Texto y de la Web



**DECSAI**

**Departamento de Ciencias de la  
Computación e Inteligencia Artificial**



Universidad de Granada

[decsai.ugr.es](http://decsai.ugr.es)

# Minería Web del Uso



**DECSAI**

**Departamento de Ciencias de la  
Computación e Inteligencia Artificial**

# Minería Web del Uso

- **Objetivo:** Comprender cómo los usuarios navegan por un sitio web o una app, qué les interesa, qué páginas o pantallas visitan, cuánto tiempo permanecen en cada una, y otros patrones de uso para mejorar la experiencia del usuario, personalizar el contenido o optimizar la estructura del sitio web o app.
  - *Datos de navegación:* Datos que registran las visitas a una página, clics, tiempo de permanencia, etc.
  - *Análisis de comportamiento:* Identificar patrones de navegación y comportamientos recurrentes.
  - *Personalización:* Mejorar la experiencia del usuario en función de sus preferencias y comportamientos previos.

# Minería Web del Uso

- Análisis de los archivos *logs* de datos de acceso de usuarios a un determinado servidor
- Tienes fines de marketing, generalmente
- Se extraen transacciones de los ficheros y se obtienen reglas de asociación o secuencias
- También se puede utilizar clustering
- Software para Minería web del uso: Google Analytics, Open Web Analytics (OWA), Woopra, Splunk, Storm, Sumo Logic, Loggly, Paper Trails, LogStash, GrayLog2, ...

# Minería Web del Uso



Servidor Web  
Servidor Proxy  
Máquina del Usuario

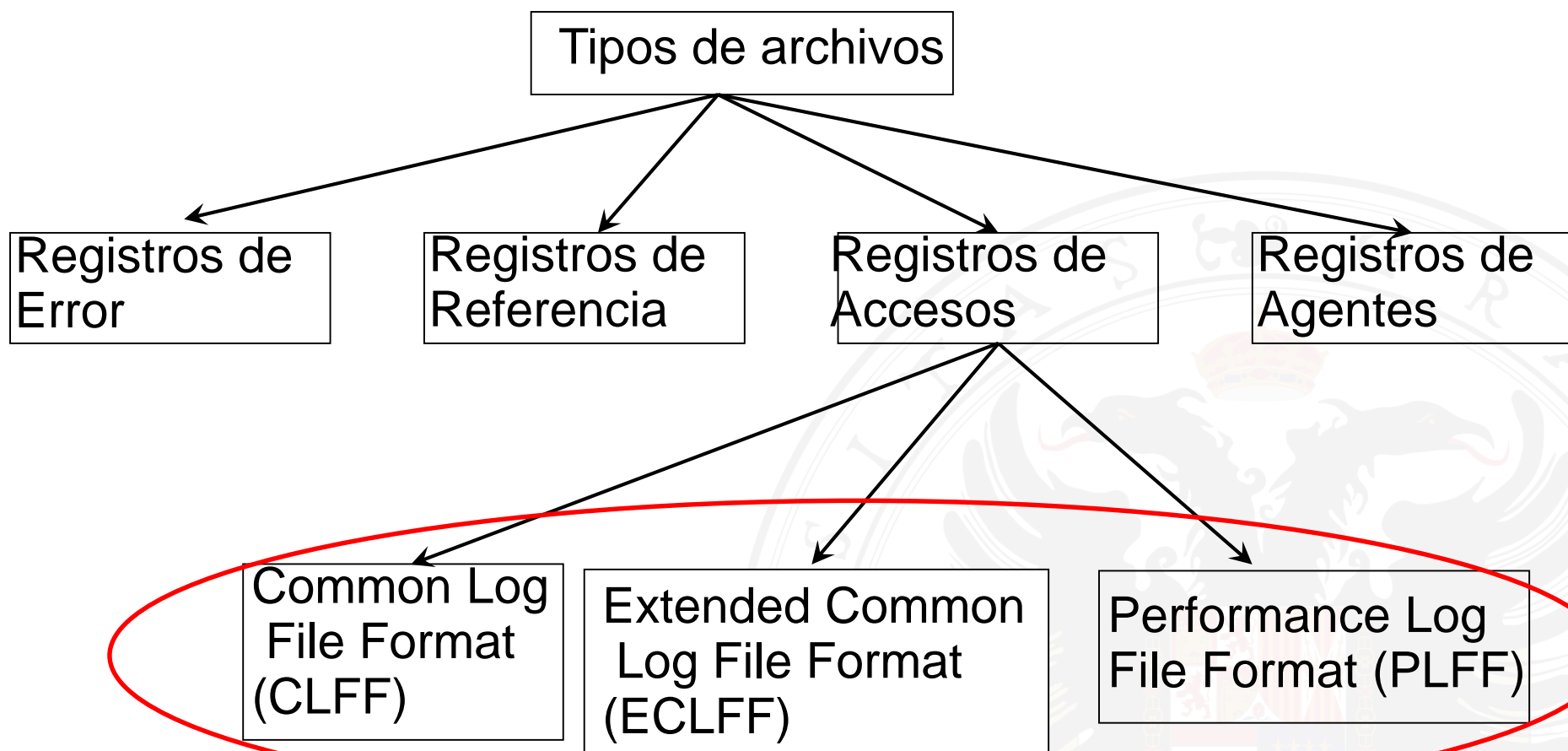
Heterogeneidad y Carencia de Estructura  
Forma Intermedia

Técnicas de Minería

- Clustering
- Reglas de Asociación

Filtrar reglas o patrones  
Construcción de perfiles

# Ficheros de log





# Ficheros de log

## *Common Log File Format (CLFF)*

Host o IP	Identificación Usuario	Autenticación Usuario	Fecha / Hora	Petición	Estado	Byte
maquina.uji.es	-	-	[9/Feb/2025:00:56:56 +0100]	"GET /documento.html HTTP/1.0"	302	64

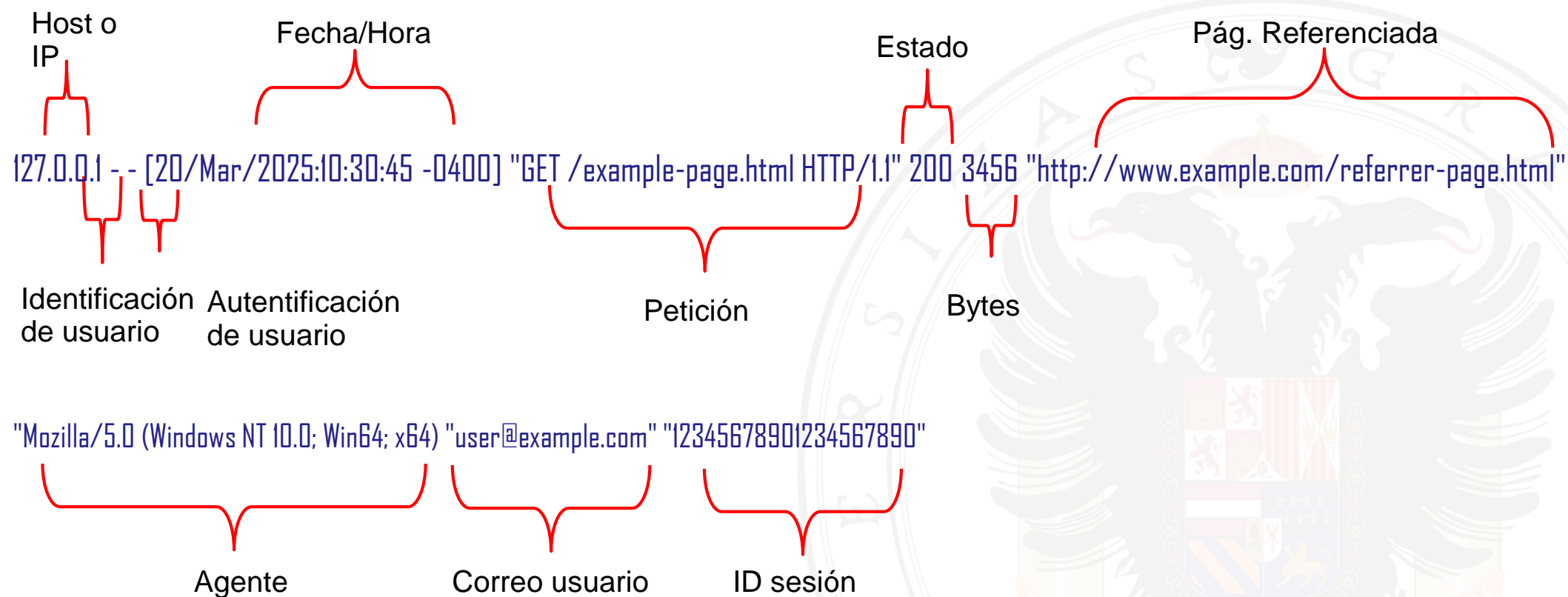
## *Extended Common Log File Format (ECLLF)*

Host o IP	Id. Usuario	Aut. Usuario	Fecha / Hora	Petición	Estado	Byte	Referencia	Agente
maquina.uji.es	-	-	[9/Feb/2025:00:56:56 +0100]	"GET /documento.html HTTP/1.0"	302	64	http://www.skyweb.com	"Mozilla/46.0.1 (Win7; I)"



# Ficheros de log

## ➤ Archivo Extended Common Log File Format (ECLFF)



# Ficheros de log

```
127.0.0.1 - - [20/Mar/2024:10:30:45 -0400] "GET /example-page.html HTTP/1.1" 200 3456  
"http://www.example.com/referrer-page.html" "Mozilla/5.0 (Windows NT 10.0; Win64; x64) "user@example.com"  
"12345678901234567890"
```

“127.0.0.1”: Es la dirección IP del cliente que realizó la solicitud.

“-”: El guion indica que la identidad del cliente no está disponible.

“-”: El guion indica que el usuario autenticado no está disponible.

“[20/Mar/2024:10:30:45 -0400]”: La fecha y hora de la solicitud en formato de registro.

“GET /example-page.html HTTP/1.1”: El método de solicitud HTTP, la URL solicitada y la versión del protocolo.

200: El código de estado de la respuesta del servidor (en este caso, 200 es "OK").

3456: El tamaño de la respuesta en bytes.

“http://www.example.com/referrer-page.html”: La URL de la página de referencia desde la que se originó la solicitud.

“Mozilla/5.0 (Windows NT 10.0; Win64; x64) “: La cadena de agente de usuario que identifica el navegador web del cliente.

“user@example.com”: Información adicional, como la dirección de correo electrónico del usuario (este campo podría ser parte del formato extendido).

“12345678901234567890”: Otro campo adicional, por ejemplo, un ID de sesión o cualquier otro dato relevante.

# Modelo de datos

- Una página web  $p_i \in P$ ,  $P = \{p_1, p_2, \dots, p_n\}$  es un documento HTML identificable a través de la red con una URL.
- Una sesión de usuario  $s_j \in S = \{s_1, s_2, \dots, s_m\}$  se define como una secuencia de accesos temporales a un sitio particular de la Web por un usuario. Viene determinada por un conjunto de clicks  $R = \{r_1, \dots, r_q\}$  en el sitio web, denominado clickstream.
- La relación entre una página  $p_i \in P$  y la sesión  $s_j \in S$  en la que se visita viene dada por la matriz  $UP$ , donde cada valor representa el tiempo de la permanencia del usuario en la página  $p_i$  en la sesión  $s_j$

$$UP = [w(s_i, p_j)], 1 \leq i \leq m, 1 \leq j \leq n$$

# Modelo de datos (cont.)

- Caracterizaremos el clickstream para la identificación de sesiones de usuarios (Método timeout [Chen et al., 1996]):
  - Sea  $r_k$  el  $k^{\text{th}}$  click del clickstream  $R$  de una dirección IP en  $t_k$  segundos.
  - Sea  $r_{k+1}$  el  $(k+1)^{\text{th}}$  click del clickstream de la misma dirección IP en  $t_{k+1}$  segundos después del click  $r_k$ .
  - $T$  es el tiempo de espera calculada como la diferencia entre ambos click en un sitio web es  $T = (t_{k+1} - t_k)$
  - Si  $T < \beta$ , siendo  $\beta$  el tiempo de espera máximo, entonces el click  $r_k$  y  $r_{k+1}$  son considerados partes de la sesión  $S_i$ . En otro caso, si  $T > \beta$ , entonces el click  $r_k$  es estimada como final de la sesión  $S_i$ , mientras el click  $r_{k+1}$  es el clickstream de la sesión  $S_{i+1}$ .

# Modelo de datos (cont.)

- Identificación de sesiones de usuario:

IP o Host	Id sesión	Fecha/Hora	Tiempo
33.red-83-33-.dynamicip.rimade.net	1	[12/Feb/2024:07:41:14]	0
1j2591.inktomisearch.com	2	[12/Feb/2024:07:41:20]	0
70.42.51.20	3	[12/Feb/2024:07:41:35]	0
33.red-83-33-8.dynamicip.rima-tde.net	1	[12/Feb/2024:07:41:39]	25

clickstream	Click
/alumnos/mlii/prolog	1
/download/guia	1
/proyectos/silviaacid/basd	1
/alumnos/oscp/fecha	2

# Patrones de navegación

- Objetivo

Obtener patrones de navegación del usuario y así tener una mejor descripción de su comportamiento en la web y de esa manera saber realmente lo que sucede en el sitio web.

- Técnicas

- Reglas de asociación
- Clustering
- Minería de secuencias

# Patrones de navegación:

## Reglas de asociación difusas

- Problema a estudiar:
  - Problema de navegación temporal:  
*intervalo horario – página visitada*
  - Problema de navegación intrapágina:  
*pagina visitada - página referenciada.*



# Patrones de navegación: Reglas de asociación difusas

- Definición: Dado  $I$  un conjunto de ítems, definiremos una transacción difusa  $\tilde{\tau}$ , donde  $\tilde{\tau} \subseteq I$ . Sea  $\tilde{\tau}(i)$  el grado de pertenencia de  $i$  a  $\tilde{\tau}$  y notaremos a  $\tilde{\tau}(I_0)$  el grado de inclusión de un ítemset en una transacción difusa definida como

$$\tilde{\tau}(I_0) = \min_{i \in I_0} \tilde{\tau}(i)$$

- Medidas de interés: hemos utilizado diferentes medidas para la obtención de las reglas de asociación difusas, tanto medidas objetivas como subjetivas
- Algoritmo: AprioriTID

# Modelo asociado a la navegación temporal - Fecha-Página visitada

- Problema 1: problema de navegación temporal.
  - Ítems: fecha/hora y páginas visitadas
  - Transacciones: tablas transaccionales difusa para la obtención de las reglas, para los ítems fecha/hora y páginas visitadas.

# Modelo asociado a la navegación temporal - **Fecha-Página visitada**

Hora.	Peso	Etiqueta
08:30	1.0	Mañana
12:45	0.5	Medio Día
15:25	0.4	Tarde
20:20	0.3	Noche

IP/Pag.	Madrugada	Mañana	Mediodía	Tarde	Noche	Pag <sub>1</sub>	Pag <sub>2</sub>	Pag <sub>3</sub>
IP <sub>1</sub>	0	1.0	0	0	0	0.4	0	0.8
IP <sub>2</sub>	0	0	0.5	0	0	0	0	0.4
IP <sub>3</sub>	0	0	0	0.4	0	0.7	0.3	0
IP <sub>4</sub>	0	0	0	0	0.3	0.2	0	0

# Modelo asociado a la navegación temporal

## Fecha - Página visitada

**Fecha/Hora → Página Visitada.**

Mañana → <https://etsiit.ugr.es/docencia/trabajo-fin-grado>

Soporte = 60%; confianza = 1.0; FC = 1.0

**-Interpretación:** del conjunto analizado el 60% presentaba esta regla, la cual nos indica que los usuarios se conectan por la mañana a esa página.

# Modelo asociado a la navegación entre páginas

## Página visitada – Página referenciada

- Problema 2: problema de navegación entre páginas.
  - Ítems: páginas visitadas y páginas referenciadas.
  - Transacciones: tablas transaccionales difusa para la obtención de las reglas, para los ítems páginas visitadas y páginas referenciadas:

# Modelo asociado a la navegación entre páginas - **Página visitada** – **Página referenciada**

Frecuencia

IP/Pag.	Pag <sub>1</sub>	Pag <sub>2</sub>	Pag <sub>3</sub>	Pag <sub>4</sub>
IP <sub>1</sub>	0	4	0	7
IP <sub>2</sub>	7	0	8	0
IP <sub>3</sub>	6	0	2	0
IP <sub>4</sub>	0	3	0	10

Peso

IP/Pag	Pag <sub>1</sub>	Pag <sub>2</sub>	Pag <sub>3</sub>	Pag <sub>4</sub>
IP <sub>1</sub>	0	0.4	0	0.7
IP <sub>2</sub>	0.7	0	0.8	0
IP <sub>3</sub>	0.6	0	0.2	0
IP <sub>4</sub>	0	0.3	0	1

Obtención de los pesos

# Modelo asociado a la navegación entre páginas - **Página visitada – Página referenciada**

**página visitada → página referenciada**

<https://etsiit.ugr.es/docencia/trabajo-fin-grado> →  
<https://swad.ugr.es/es>

- Soporte =40%; confianza =1.0; FC =1.0

**Interpretación:** esto indica que los usuarios visitan la página <https://etsiit.ugr.es/docencia/trabajo-fin-grado> y luego se van a la página <https://swad.ugr.es/es>. Esta regla se encuentra en un 40% dentro del conjunto analizado.





Universidad de Granada

[decsai.ugr.es](http://decsai.ugr.es)

# Minería Web de uso: clustering y perfiles de usuario



**DECSAI**

**Departamento de Ciencias de la  
Computación e Inteligencia Artificial**

# Contenidos

- Tecnología de minería web de uso en un portal web
- Clustering de sesiones de usuario
- Definición de perfiles de usuario en XML
- Obtención de perfiles de usuario a partir del clustering

# Usuarios de negocios/marketing

- Los usuarios de negocios en las empresas son los que tratan con la información extraída de los procesos de minería
  - Identificación de grupos sociales
  - Relación entre los usuarios (clientes) y sus áreas de interés
- Mala comunicación entre los analistas de minería de datos y los usuarios de negocios
- Necesidad de software específico para el desarrollo de herramientas de minería para usuarios de negocios

# Perfiles de usuario en la web

- Se almacena información sobre el comportamiento de los usuarios al navegar
  - Información de minería para el usuario
  - Información de minería para el usuario de negocios o de marketing

# Perfiles de usuario en la web

## Información de minería para el usuario

El sitios web comerciales, la empresa puede extraer las preferencias del usuario y personalizar el sitio web, ofreciendo al usuario algunos productos de acuerdo a sus preferencias.

# Perfiles de usuario en la web

- Información de minería para usuarios de negocios o de marketing
  - Proceso de Clustering para agrupar perfiles de usuarios por áreas de interés
  - Se puede generar un conjunto de reglas para realizar inferencia de la relación entre usuarios y términos
  - Conexión de usuarios a grupos sociales basados en información extraída de los meta-datos de la página web

# Usuarios registrados y no registrados

- ¿Identidad del usuario?
- Los usuarios con la misma dirección IP tienen diferentes perfiles
- Los usuarios pueden proporcionar nicknames



# Usuarios registrados y no registrados

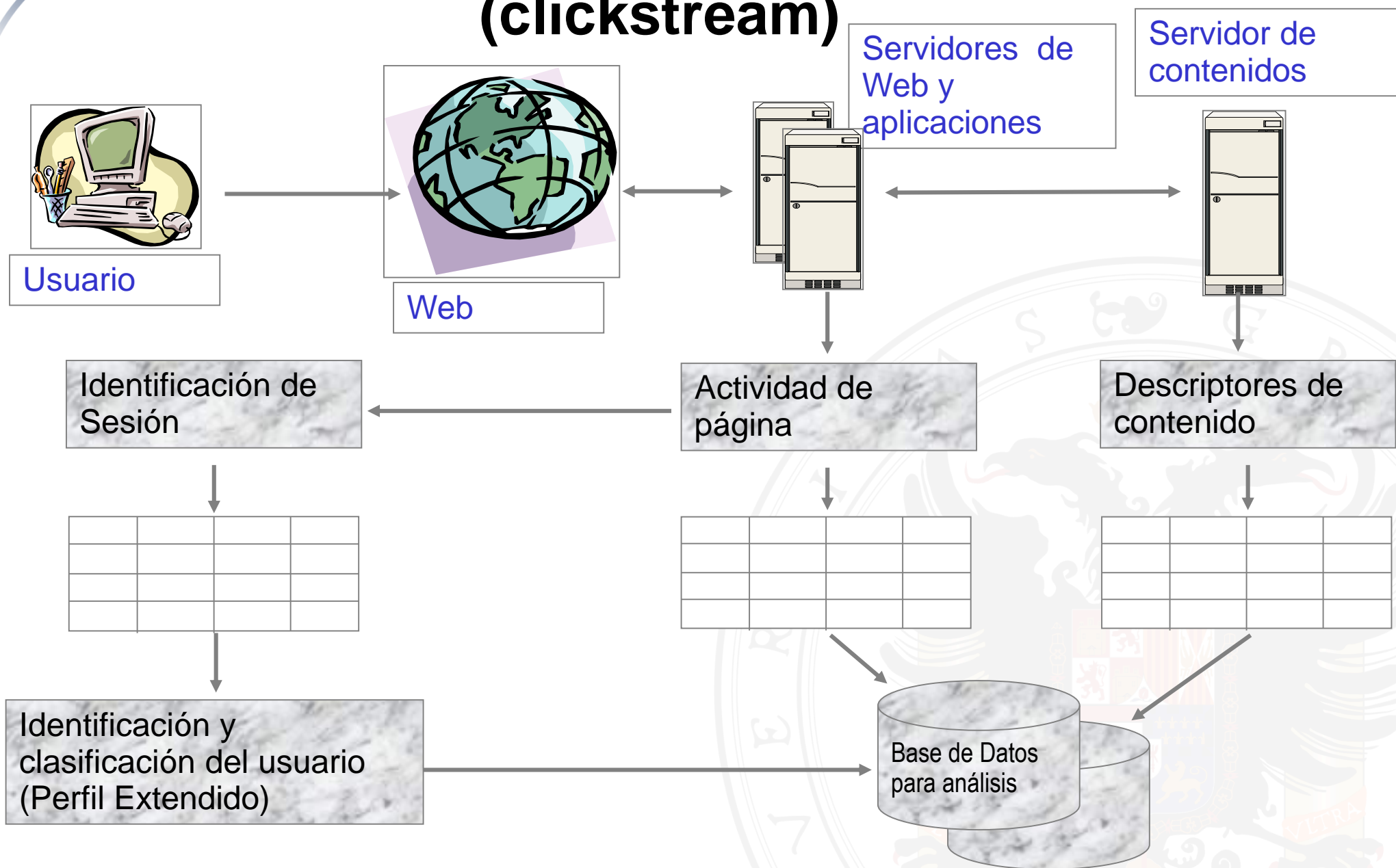
Usuarios no registrados

Personalización

Usuarios registrados

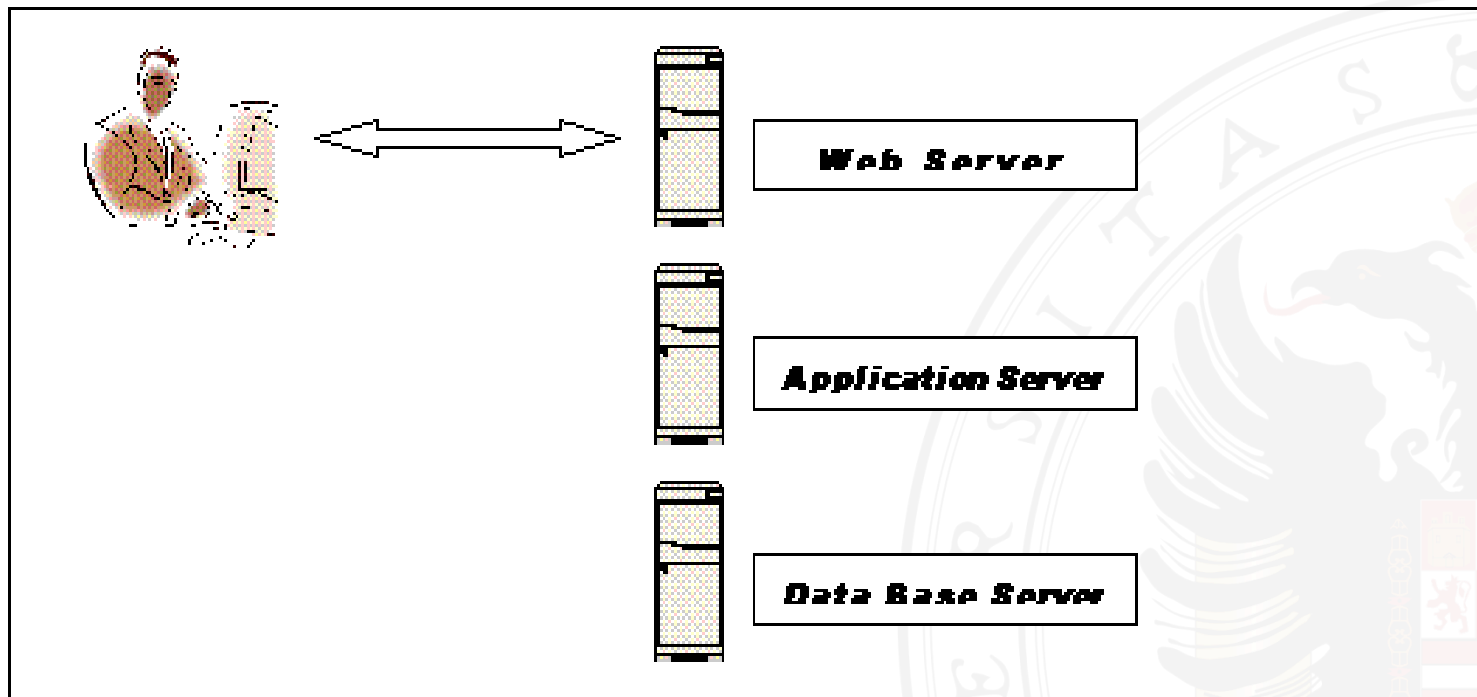
Customización

# Procesamiento de la actividad de ratón (clickstream)



# Tecnología de minería en Web en un portal

- El área del portal



- El área de análisis de datos

# Tecnología de minería en Web en un portal

- El área de análisis de datos
  - Entrada de datos:  
Explícita: Formularios, Opiniones, registros,...  
Implícita: Cookies, ficheros de log
  - Generación de datos:  
Ficheros de logs  
Ficheros de metadatos
  - Filtrado de datos:  
Eliminar datos irrelevantes o erróneos (p.ej. Bots o accesos no humanos)
- Objetivo.  
Obtener grupos de sesiones de usuarios que navegan por la web con características similares.

# Clustering en minería web de uso

- **Objetivo:**  
Obtener grupos de sesiones de usuarios que navegan por la web con características similares.
- **Modelo de datos asociado:**
  - Conjunto de sesiones  $S$  definida como:  $S = \{s_1, s_2, \dots, s_m\}$
  - Conjunto de páginas  $P$ , definida como:  $P = \{p_1, p_2, \dots, p_n\}$
  - Matriz sesión-página  $m \times n$ :  $UP = [w(s_i, p_j)], 1 \leq i \leq m, 1 \leq j \leq n$
- A partir de esta matriz de peso sesión-página obtendremos una matriz de semejanza entre las sesiones definida

$$SS = [sim(s_i, s_j)], i \leq j \leq m$$

y para obtener esta semejanza aplicaremos la medida del coseno y coseno extendido.

# Medidas de semejanza

- Coseno. 
$$S_{1,kl} = \frac{\sum_{i=1}^N \sum_{j=1}^N S_i^k \cdot S_j^l}{\sqrt{\sum_{i=1}^N S_i^k} \sqrt{\sum_{j=1}^N S_j^l}}$$

- Coseno extendido. 
$$S_{2,kl} = \frac{\sum_{i=1}^N \sum_{j=1}^N S_i^k S_j^l Sn(i, j)}{\sqrt{\sum_{i=1}^N S_i^k} \sqrt{\sum_{j=1}^N S_j^l}}$$

$$Sn(i, j) = \min \left( \frac{|p_i \cap p_j|}{\max(1, \max(|p_i|, |p_j|) - 1)} \right)$$

# Ejemplo de clustering en minería web de uso

➤ Conjuntos de datos:

Conjuntos de datos	Entrada de datos originales	Entrada de datos preprocesadas	Nº Sesiones
Conjunto 1	100900	12910	2024
Conjunto 2	98202	15676	2780

- Medidas: coseno y coseno extendido.
- Técnica: clustering difuso con el algoritmo c-medias difuso.
- Número de particiones iniciales: 12 (obtenidos mediante el análisis jerárquico previamente)



# Ejemplo de Clustering en minería web de uso

Nº Cluster	Sesiones (Grado de Pertenencia)	Sesión Centroide
Cluster 0	2 (0.95)	GET/apps/tablon
	437 (0.98)	GET/apps/foro/index.php
	508 (0.98)	GET/apps/foro/index.php?action=foro&idforo=escuela
	512 (0.96)	GET/apps/foro/index.php?action=foro&idforo=general
		GET/apps/foro/index.php?action=hebra&idhebra=1920
		GET/apps/foro/index.php?action=foro&idforo=asignaturas
		GET/apps/foro/index.php?action=hebra&idhebra=1937
		GET/apps/foro/index.php?action=hebra&idhebra=1920
Cluster 3	21 (1.00)	GET/apps/foro/index.php?action=hebra&idhebra=1916H
	65 (1.00)	
	6 (1.00)	GET/js/protWindows/themes/default.css
	51(1.00)	GET/apps/foro/index.php
	136 (1.00)	GET/apps/tablon
	13 (1.00)	GET/page.php?pageid=departamentos
	68 (0.939)	GET/apps/foro/index.php?action=hebra&idhebra=1583
	569 (0.939)	GET/apps/foro/index.php?action=hebra&idhebra=1874
		GET/apps/foro/index.php?action=foro&idforo=escuela
		GET/apps/foro/index.php?action=hebra&idhebra=1709
		GET/apps/foro/index.php?action=foro&idforo=general

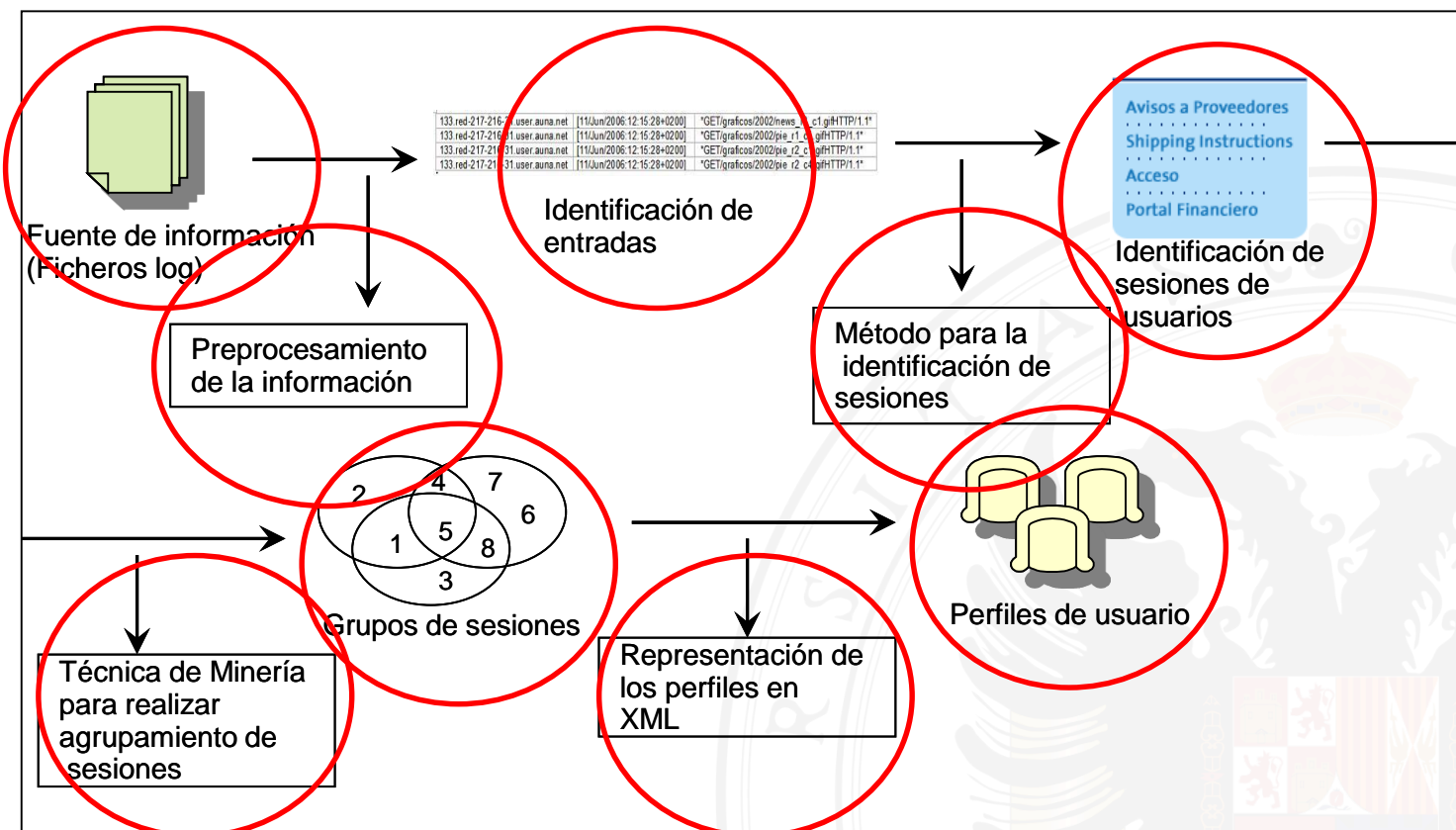
Coseno extendido

Nº Cluster	Sesiones (Grado de Pertenencia)	Sesión Centroide
Cluster 0	2 (0.75)	GET/apps/tablon
	437 (0.90)	GET/apps/foro/index.php
	508 (0.90)	GET/apps/foro/index.php?action=foro&idforo=escuela
	512 (0.85)	GET/apps/foro/index.php?action=foro&idforo=general
		GET/apps/foro/index.php?action=hebra&idhebra=1920
		GET/apps/foro/index.php?action=foro&idforo=asignaturas
		GET/apps/foro/index.php?action=hebra&idhebra=1937
		GET/apps/foro/index.php?action=hebra&idhebra=1920
Cluster 3	21 (0.97)	GET/apps/foro/index.php?action=hebra&idhebra=1916H
	65 (0.97)	
	6 (0.97)	GET/js/protWindows/themes/default.css
	51(0.97)	GET/apps/foro/index.php
	136 (0.97)	GET/apps/tablon
	13 (0.97)	GET/page.php?pageid=departamentos
	68 (0.85)	GET/apps/foro/index.php?action=hebra&idhebra=1583
	569 (0.85)	GET/apps/foro/index.php?action=hebra&idhebra=1874
		GET/apps/foro/index.php?action=foro&idforo=escuela
		GET/apps/foro/index.php?action=hebra&idhebra=1709
		GET/apps/foro/index.php?action=foro&idforo=general

Coseno

- Discusión de los resultados: la medida del coseno extendido da mejores resultados con respecto a los centroides

# Obtención de perfiles de usuario en minería web de uso



# Minería de secuencias

**Objetivo:** identificar las secuencias de acciones que un usuario realiza en un sitio web o en una aplicación móvil, con el objetivo de entender cómo interactúan con la interfaz y qué rutas siguen al navegar.

Este análisis permite a las empresas optimizar la experiencia del usuario, mejorar la usabilidad, personalizar contenidos y hacer recomendaciones más precisas.

# Minería de secuencias

La minería de secuencias se aplica principalmente en:

- **Análisis de comportamiento del usuario:** Identificación de patrones comunes en las rutas de navegación.
- **Optimización de la experiencia del usuario:** Descubrimiento de posibles mejoras en la estructura del sitio o la aplicación para reducir la tasa de rebote.
- **Recomendaciones personalizadas:** A partir de las secuencias de navegación, se pueden sugerir productos, páginas o servicios que podrían interesar al usuario.
- **Prevención de abandono:** Detectar los puntos donde los usuarios tienden a abandonar la sesión y mejorar esos momentos.

# Algoritmos de minería de secuencias

- **Algoritmo GSP (Generalized Sequential Pattern)** Srikant, R., & Agrawal, R. (1996).

El **GSP** es una extensión del algoritmo Apriori, diseñado específicamente para la minería de patrones secuenciales.

- **Algoritmo SPADE (Sequential PAttern Discovery using Equivalence classes)** Zaki, M. J. (2001).

**SPADE** es un algoritmo diseñado para detectar patrones secuenciales utilizando una representación eficiente de las secuencias mediante **clases de equivalencia**. Su principal ventaja es que puede manejar grandes volúmenes de datos de manera más eficiente que otros algoritmos.

- **Algoritmo CloSpan (Closed Sequential Pattern Mining)** Pei, J., Han, J., & Yu, Y. (2000).

**CloSpan** está diseñado para encontrar **patrones secuenciales cerrados**, es decir, patrones que no tienen superconjuntos con la misma frecuencia. Este algoritmo mejora la eficiencia en la minería de patrones secuenciales al centrarse solo en los patrones más relevantes. Su implementación es más compleja

# Ejemplo de Minería de Secuencias

## **Ejemplo Práctico: Análisis de la Navegación de Usuarios en un E-commerce**

Imaginemos que tenemos un sitio web de e-commerce y queremos analizar cómo los usuarios navegan por el sitio para mejorar la conversión de ventas. Para ello, aplicamos minería de secuencias utilizando el algoritmo GSP.

### **Paso 1: Recolección de Datos**

En primer lugar, recopilamos las secuencias de navegación de los usuarios. Cada secuencia representa una serie de páginas visitadas, con el siguiente formato:

Usuario 1: Página principal → Categoría de productos → Producto A → Carrito → Pago.

Usuario 2: Página principal → Categoría de productos → Producto B → Carrito → Pago.

Usuario 3: Página principal → Categoría de productos → Abandono.



# Ejemplo de Minería de Secuencias

## Paso 2: Aplicación del Algoritmo GSP

Utilizando el algoritmo GSP, encontramos patrones frecuentes entre las secuencias de los usuarios. Por ejemplo, podemos identificar que:

**Patrón 1:** Los usuarios que visitan la "Categoría de productos" y luego agregan un producto al carrito son más propensos a completar la compra.

**Patrón 2:** Los usuarios que abandonan el carrito después de visitar la "Categoría de productos" tienden a abandonar en un porcentaje más alto que aquellos que ya tienen un Producto en su carrito.

# Ejemplo de Minería de Secuencias

## Paso 3: Optimización de la Navegación

Con estos patrones, podemos tomar decisiones informadas sobre cómo mejorar la estructura del sitio:

**Recomendación 1:** Colocar productos más relevantes en la página de inicio o en la categoría de productos para fomentar la conversión.

**Recomendación 2:** Implementar recordatorios o descuentos a los usuarios que abandonan, basándonos en el patrón de abandono observado.





Universidad de Granada

[decsai.ugr.es](http://decsai.ugr.es)

# Visualización en Minería de Textos y de Web



**DECSAI**

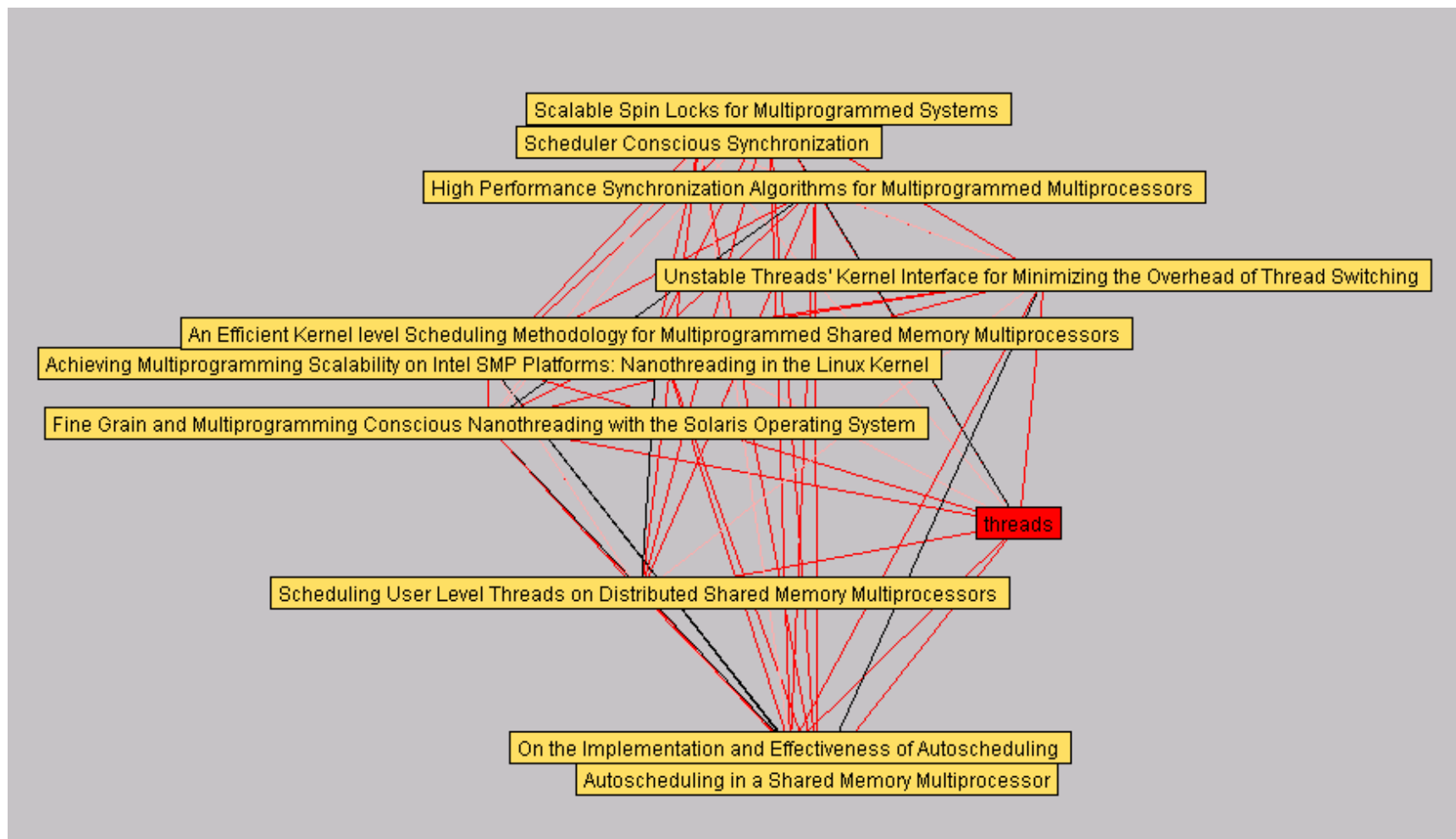
**Departamento de Ciencias de la  
Computación e Inteligencia Artificial**

# Visualización

- Clustering
- Reglas de asociación
- Tag Clouds

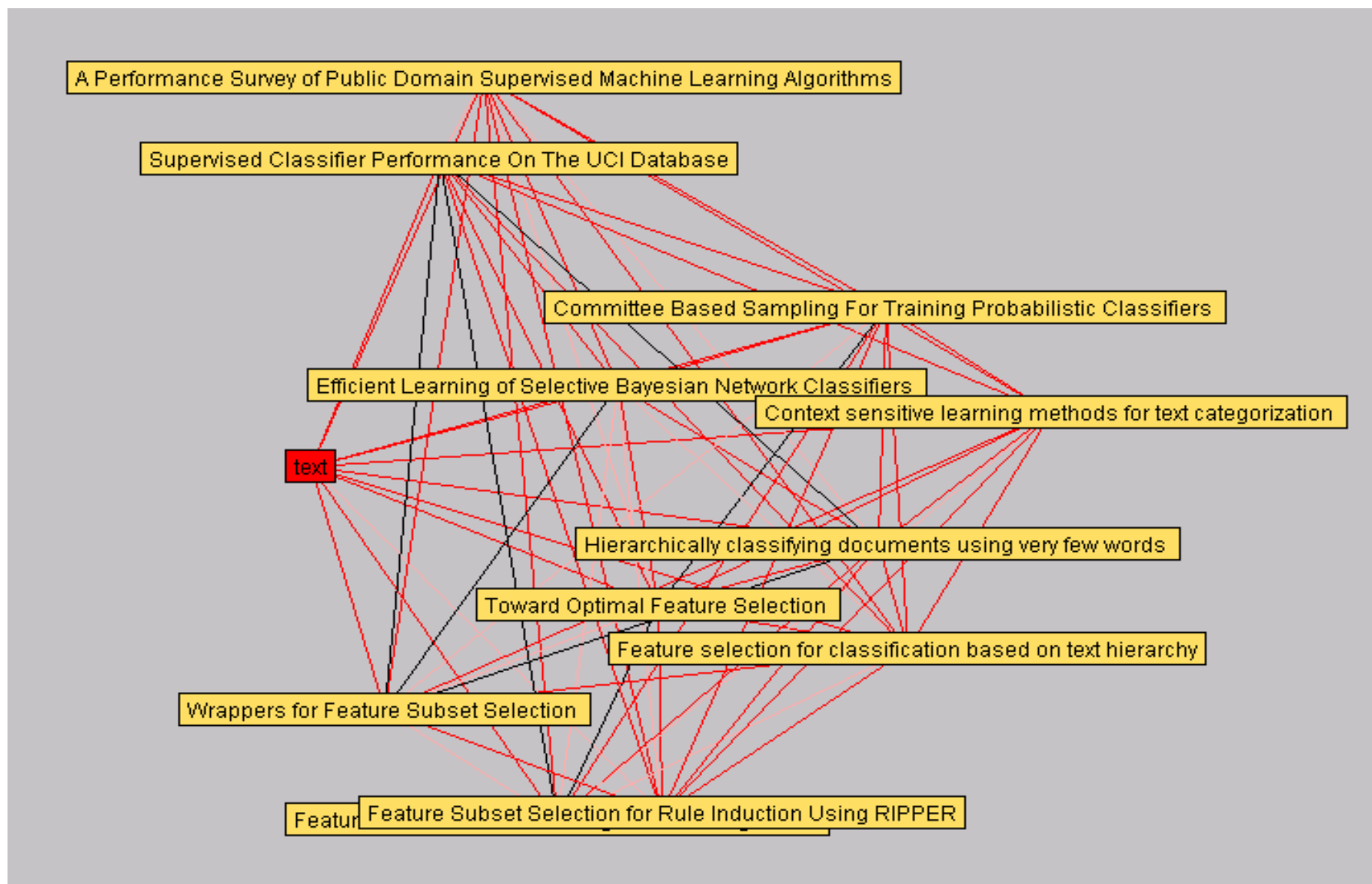


# Visualización de clusters



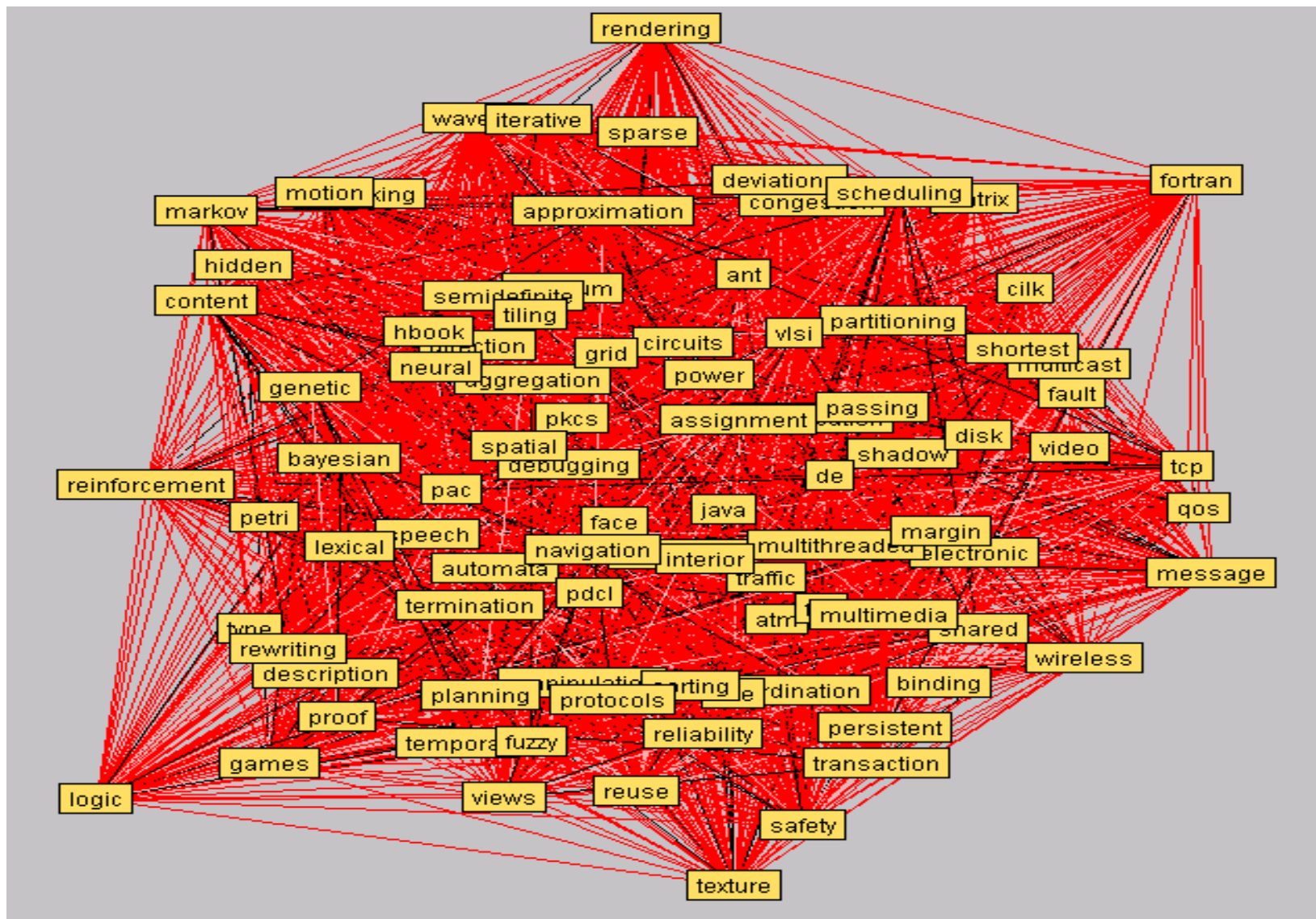
Fuente: <http://nlp.stanford.edu/courses/cs224n/2003/fp/millersj/cs224nfp.pdf>

# Visualización de clusters



Fuente: <http://nlp.stanford.edu/courses/cs224n/2003/fp/millersj/cs224nfp.pdf>

# Visualización de clusters



Fuente: <http://nlp.stanford.edu/courses/cs224n/2003/fp/millersj/cs224nfp.pdf>

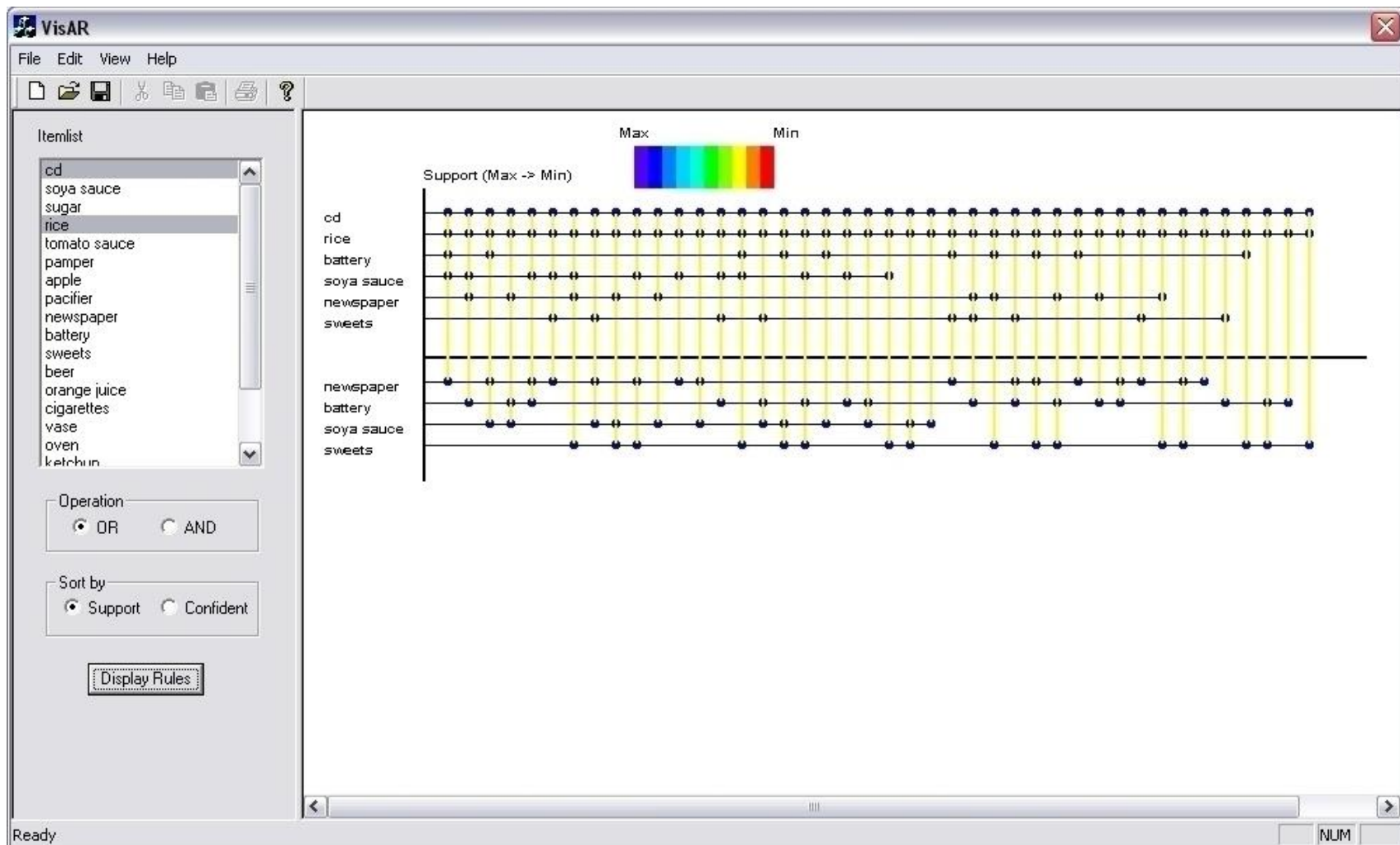
# Visualización de clusters

- Ejemplos:
- Visualizador de Clusters de Tweets

<https://www.csc2.ncsu.edu/faculty/healey/social-media-viz/production/>

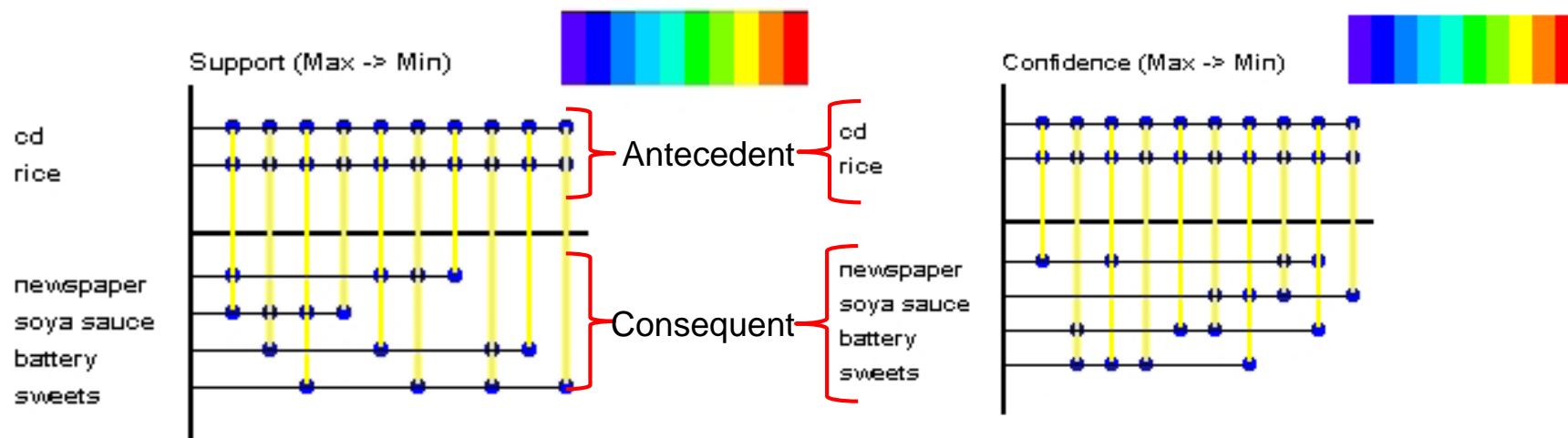


# Visualización de reglas de asociación



Fuente: K. Techapichetvanich and A. Datta. VisAR : A New Technique for Visualizing Mined Association Rules. X. Li, S. Wang, and Z.Y. Dong (Eds.): ADMA , LNAI 3584, pp. 88–95, 2005.

# Visualización de reglas de asociación

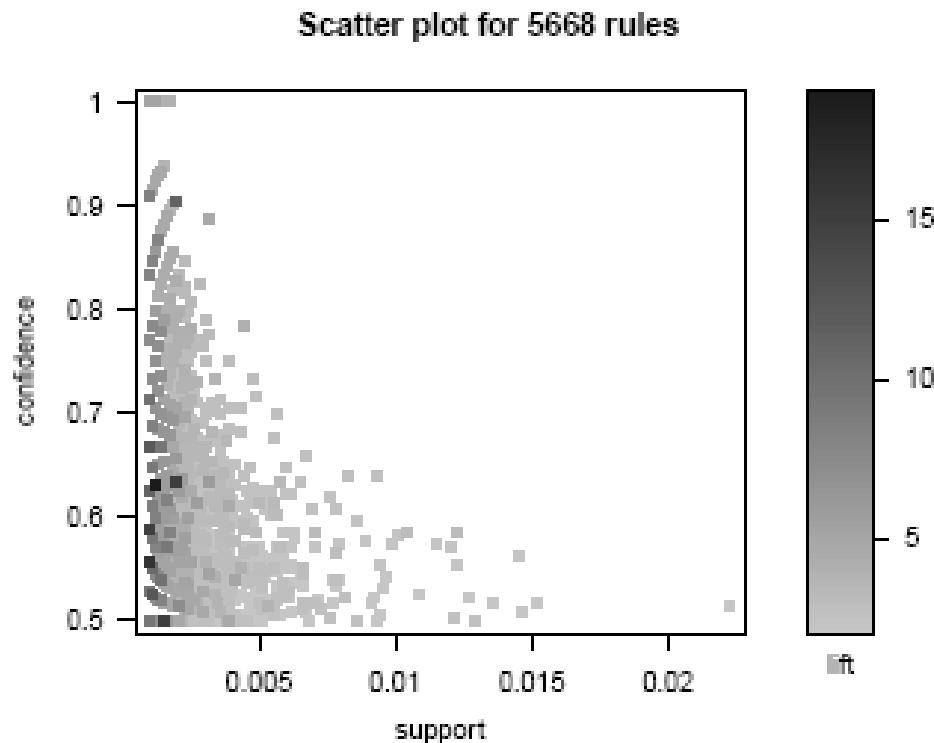


Las reglas se ordenan de izquierda a derecha ordenadas por soporte y por confianza



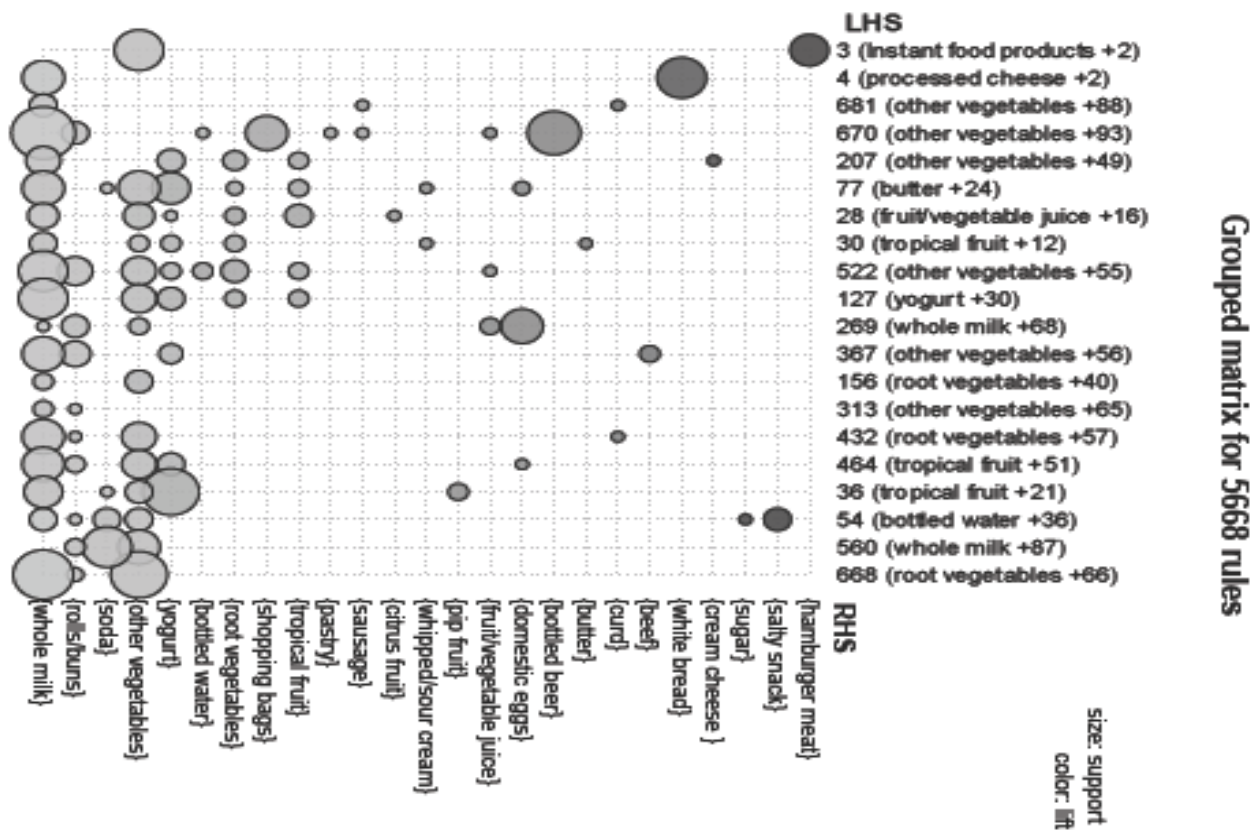
# Visualización de reglas de asociación

- Ejemplo:
  - Arulesviz Package



# Visualización de reglas de asociación

- Arulesviz Package -> reglas de asociación en clusters



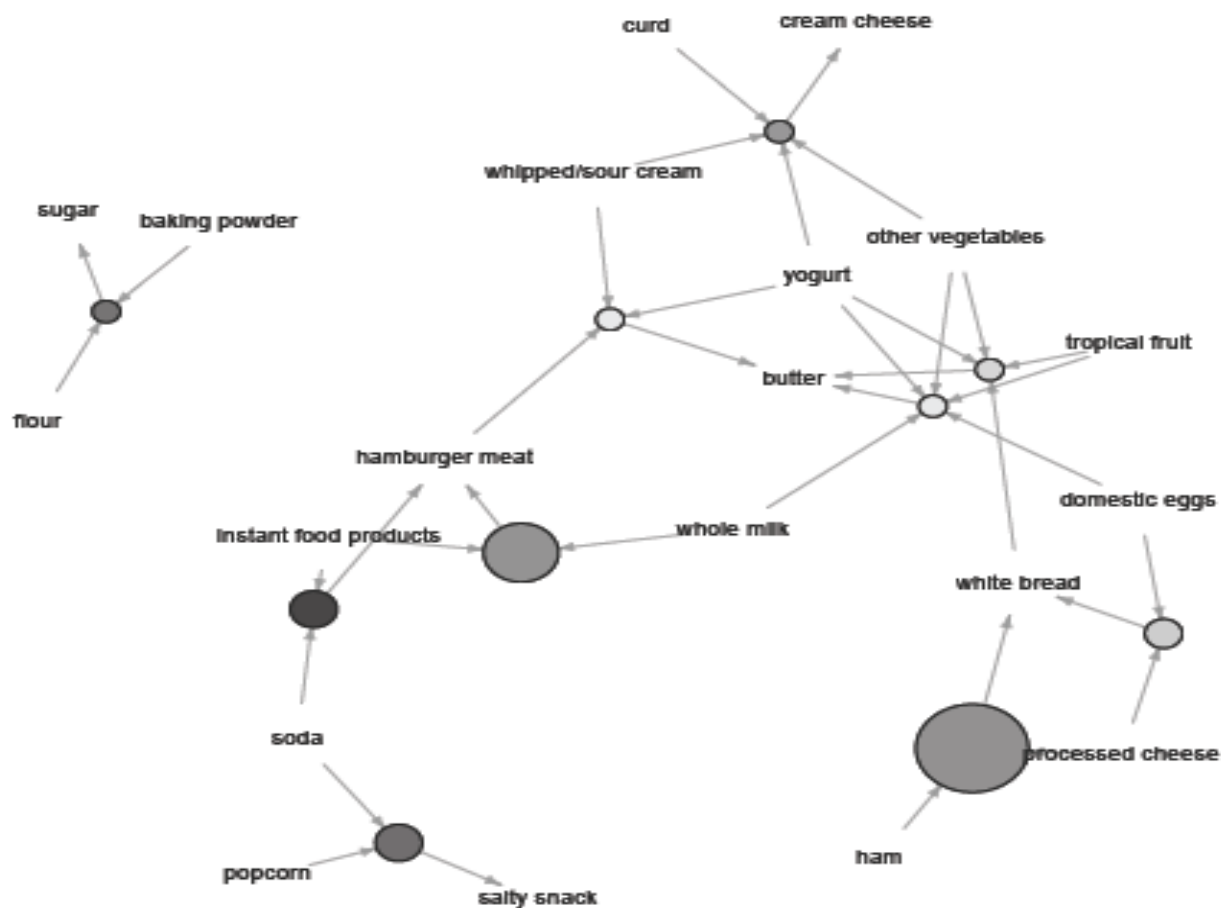
# Visualización de reglas de asociación

- Arulesviz Package (CRAN) -> reglas de asociación para R.
- <https://cran.r-project.org/web/packages/arulesViz/index.html>
- Visualización:
  - Los vértices representan items o itemsets
  - Las flechas representan relaciones en las reglas
  - El nivel de gris mide la fuerza de la asociación
  - Esta visualización no se recomienda para grupos de reglas muy grandes

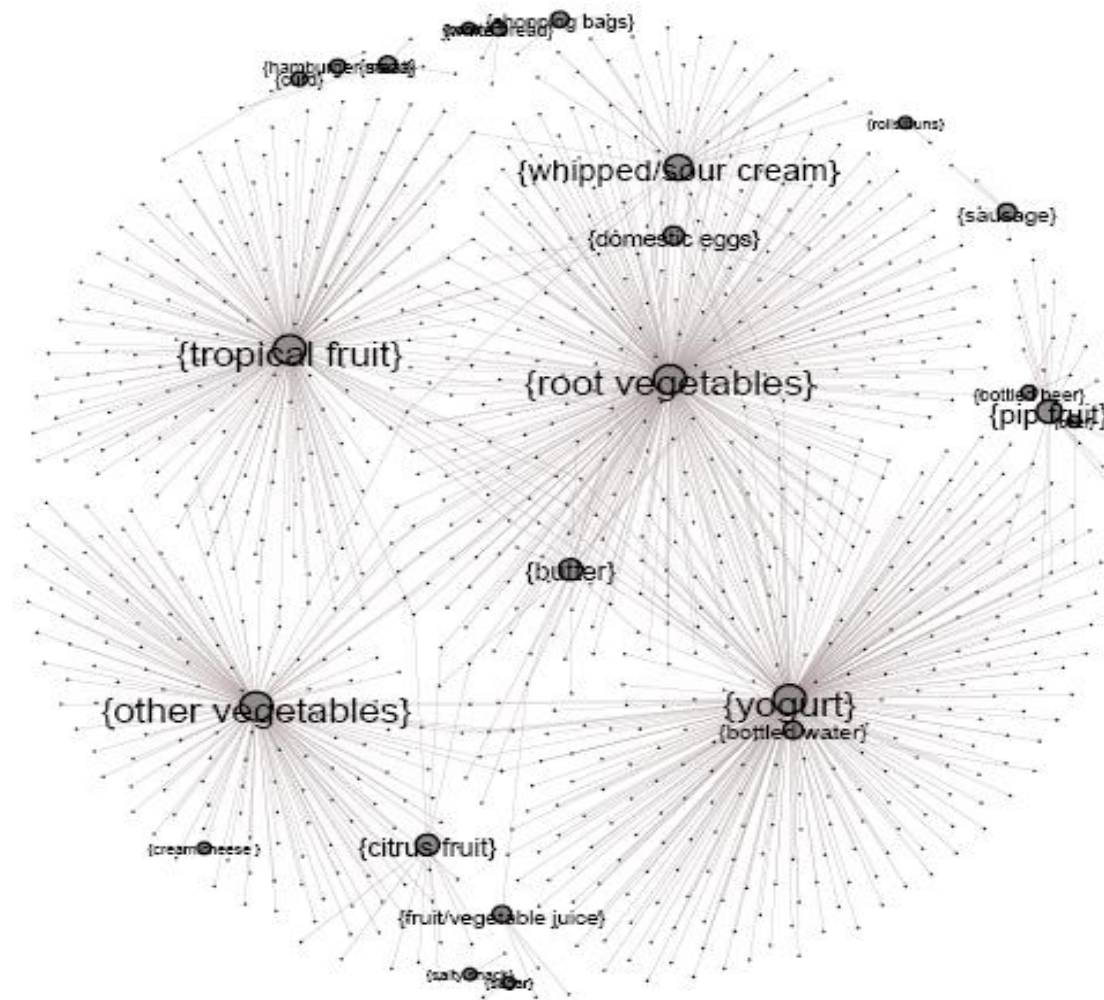
# Visualización de reglas de asociación

Graph for 10 rules

size: support (0.001 – 0.002)  
color: lift (11.279 – 18.996)



# Visualización de reglas de asociación



# Tag clouds

- Conceptualmente son similares a los histogramas
- Utilizan diferentes fuentes para representar el texto en base a:
  - Tamaño
  - Fuente
  - Orientación del texto (vertical u horizontal)
  - La proximidad de unas palabras a otras



# Tag clouds



# Tag Clouds

- Ejemplos:
  - Generador de TagClouds On-line
  - Visualizador de Tag Clouds de Tweets

<http://tagcrowd.com/>

<https://www.csc2.ncsu.edu/faculty/healey/social-media-viz/production/>