

3º Reunión Picto - Mar del Plata

Prototipo: Sistema de Sugerencias



Universidad Nacional del Sur
Biblioteca Central Prof. Nicolás Matijevic

Autores: Lic. Mitzig Mónica – Lic. Mitzig Natalia.
Abril 2013.

Indice

- Herramienta para la catalogación en RI.
- Estructura del sistema.
- Funcionamiento.
- Evaluación de la efectividad de las pruebas.
- Licencia.
- Trabajos futuros.
- Preguntas.
- Agradecimientos.
- Contacto.

Prototipo Sugerencias

Herramienta para la catalogación en Repositorios Institucionales

- Es un sistema de sugerencias de temáticas y palabras clave según texto seleccionado por el usuario.
- **Requisito** para su uso: Mozilla Firefox 10 o superior.



Prototipo Sugerencias

Estructura del prototipo:



Crawler o cosechador
(Crawler SciELO)



Clasificador de artículos en temáticas
(Temáticas SciELO)



Complemento para
Mozilla Firefox
(Sugerencias SciELO)



Motor de búsqueda
(Web Service SciELO)

Crawler o cosechador

Crawler SciELO



- Herramienta que permite cosechar las temáticas asociadas a las revistas y todos los artículos de revistas de los diferentes sitios *SciELO*.

Clasificador de temáticas

Temáticas SciELO



- Herramienta que clasifica los artículos de las revistas SciELO por temática/s.
- Cada temática se representa con:
 - ✓ La lista de palabras clave
 - ✓ La lista de títulos
 - ✓ La lista de resúmenesde los artículos de esa temática.

Clasificador de temáticas

Temáticas SciELO



- Se crea un índice de temáticas.
- Se realiza similitud entre título, palabras clave y resumen de los artículos de aquellas revistas que tienen más de una temática asociada y las temáticas.

Clasificador de temáticas

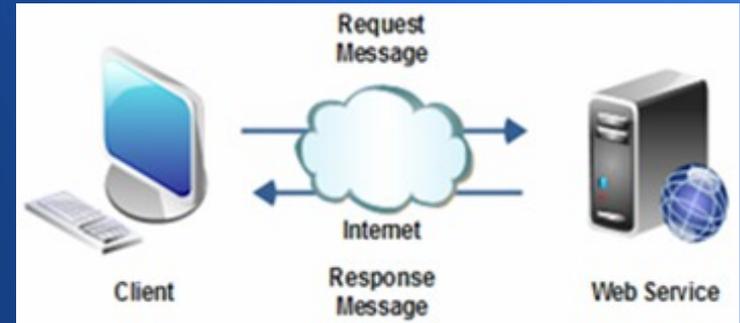
- Se asigna a cada artículo la temática de la primer sugerencia que coincida con alguna de las temáticas asociadas a la revista entre las tres mejores sugerencias obtenidas.
- Finalmente, se crea un índice de artículos que es utilizado por el servicio web.

Motor de búsqueda

Web Service SciELO



- Un ***servicio web*** permite apoyar la interoperabilidad entre máquinas que interactúan en la web.



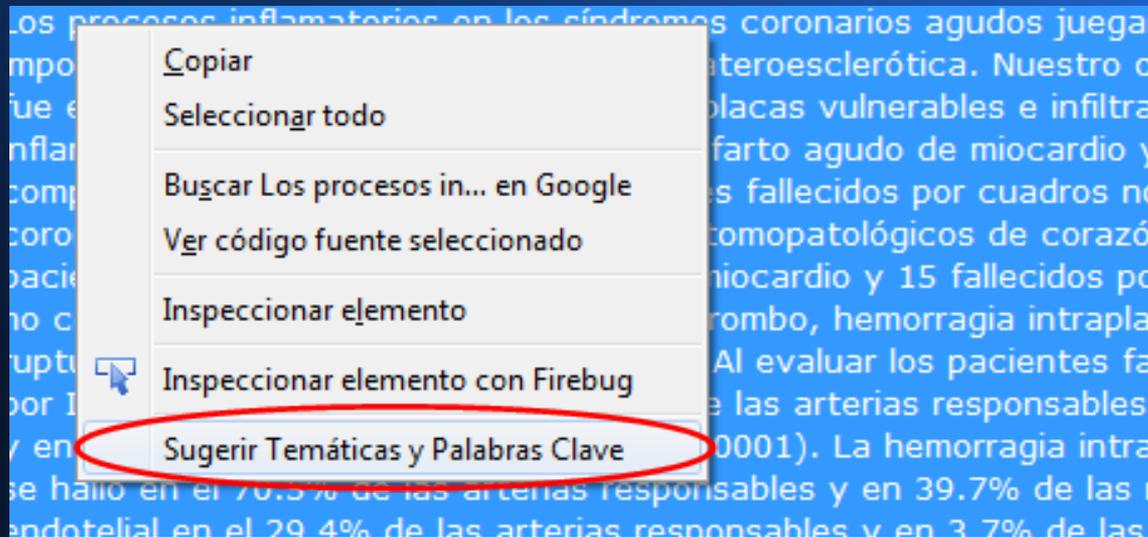
- Utiliza el índice de artículos para dar las **3 mejores sugerencias** de temáticas y palabras clave según texto dado.

Complemento Sugerencias

Sugerencias SciELO.



- Se agrega un ítem al menú contextual.



Complemento Sugerencias

- Se muestra un panel y en orden de relevancia, las **3 mejores sugerencias** de temáticas y palabras clave.

 **Sugerencias SciELO** 

Palabras relevantes: fallecidos; responsables; intraplaca; trombo; inflamatorios; pacientes; coronarios; arterias; hemorragia; miocardio; infarto; agudo; evaluar; presencia;

1° sugerencia:

Temática:
CIENCIAS DE LA SALUD

Palabras clave:
síndrome coronario agudo; infarto del miocardio con elevación del st; bloqueo de rama; bloqueo de rama derecha; mortalidad; méxico;

2° sugerencia:

3° sugerencia:

...entes responsables y en 37,7% de las no responsa

Funcionamiento

- Instalación del complemento.
- Pruebas.



Evaluación de Sugerencias

- Catalogación de recursos a partir de 45.742 artículos provenientes de 361 revistas indexadas en el repositorio SciELO.
- Herramienta automática para medir la precisión de las sugerencias obtenidas.
- Total de temáticas: 12.
- Se utilizaron 10 artículos de cada temáticas (no indexados).

Evaluación de Sugerencias

<i>Temáticas</i>	<i>Cantidad de artículos indexados</i>
T1: CIENCIAS AGRICOLAS	2403
T2: CIENCIAS BIOLÓGICAS	4250
T3: CIENCIAS DE LA SALUD	23754
T4: CIENCIAS EXACTAS Y DE LA TIERRA	2775
T5: CIENCIAS GEOLÓGICAS	197
T6: CIENCIAS SOCIALES	1705
T7: CIENCIAS SOCIALES APLICADAS	4421
T8: HUMANIDADES	4652
T9: INGENIERIA	1289
T10: LINGÜÍSTICA, LETRAS Y ARTES	164
T11: MATEMÁTICA	9
T12: QUÍMICA	123
Total de artículos	45742

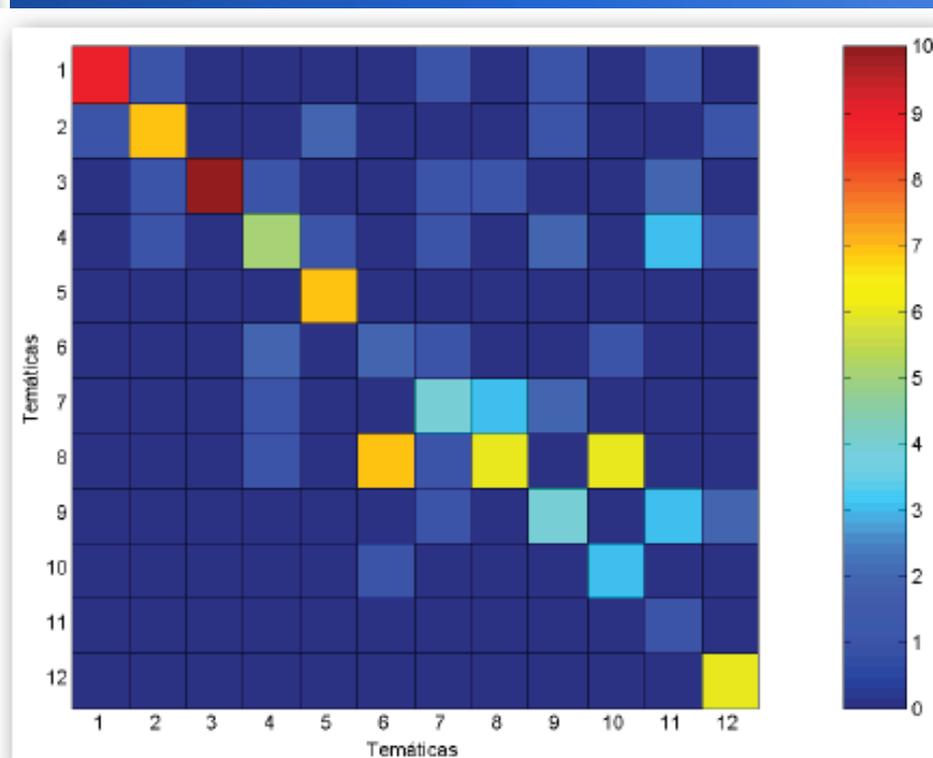
Evaluación de Sugerencias

- Para obtener las tres mejores sugerencias de los artículos de prueba, utilizamos el título, resumen y las palabras clave del mismo.
- Se evalúa la correctitud de la predicciones obtenidas, teniendo en cuenta:
 - 1) La primera predicción.
 - 2) Las dos primeras predicciones.
 - 3) Las tres predicciones.

Evaluación de Sugerencias

Matriz de confusión: primera predicción.

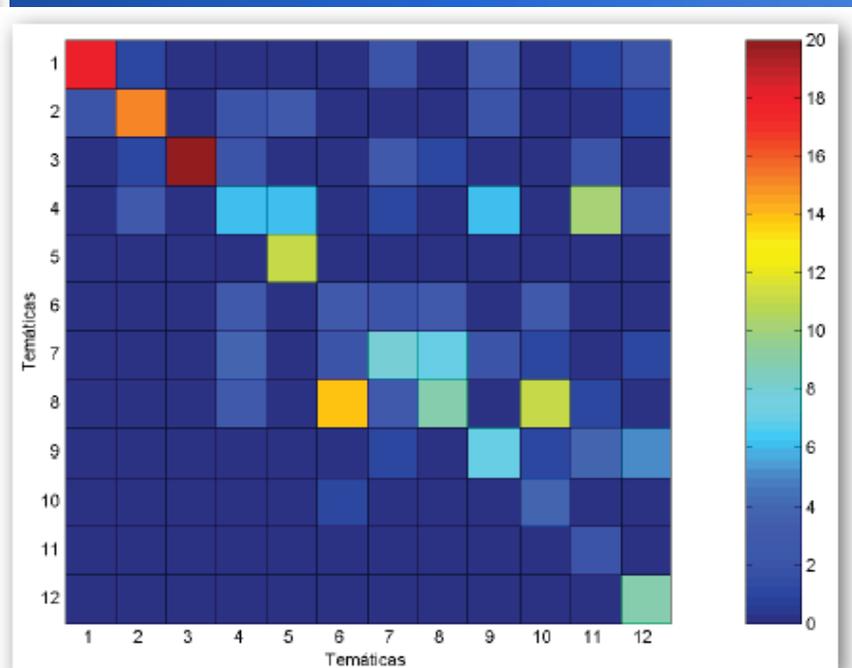
1° Sug.	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12
T1	9	1	0	0	0	0	1	0	1	0	1	0
T2	1	7	0	0	2	0	0	0	1	0	0	1
T3	0	1	10	1	0	0	1	1	0	0	2	0
T4	0	1	0	5	1	0	1	0	2	0	3	1
T5	0	0	0	0	7	0	0	0	0	0	0	0
T6	0	0	0	2	0	2	1	0	0	1	0	0
T7	0	0	0	1	0	0	4	3	2	0	0	0
T8	0	0	0	1	0	7	1	6	0	6	0	0
T9	0	0	0	0	0	0	1	0	4	0	3	2
T10	0	0	0	0	0	1	0	0	0	3	0	0
T11	0	0	0	0	0	0	0	0	0	0	1	0
T12	0	0	0	0	0	0	0	0	0	0	0	6



Evaluación de Sugerencias

Matriz de confusión: las dos primeras predicciones.

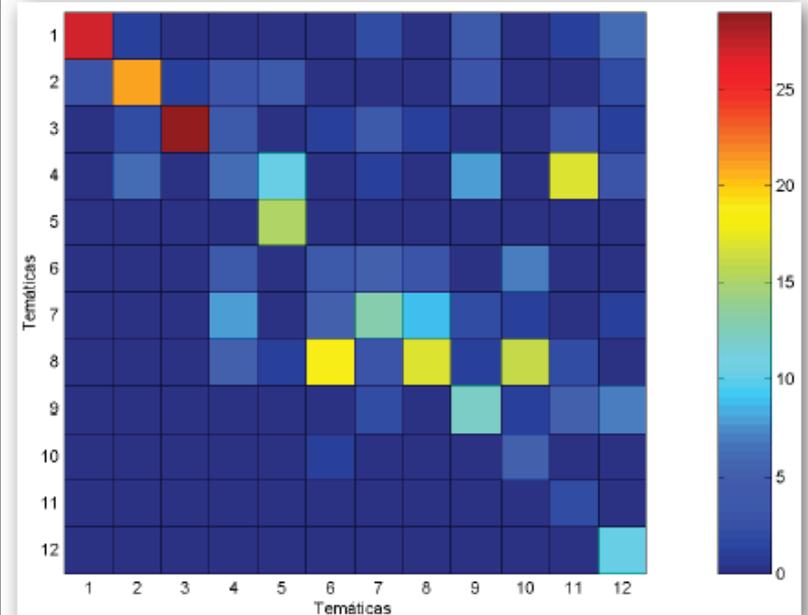
Entre las 2 Sug.	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12
T1	18	1	0	0	0	0	2	0	3	0	1	2
T2	2	15	0	2	3	0	0	0	2	0	0	1
T3	0	1	20	2	0	0	3	1	0	0	2	0
T4	0	3	0	6	6	0	1	0	6	0	10	2
T5	0	0	0	0	11	0	0	0	0	0	0	0
T6	0	0	0	3	0	3	2	3	0	3	0	0
T7	0	0	0	4	0	2	8	7	2	1	0	1
T8	0	0	0	3	0	14	3	9	0	11	1	0
T9	0	0	0	0	0	0	1	0	7	1	4	5
T10	0	0	0	0	0	1	0	0	0	4	0	0
T11	0	0	0	0	0	0	0	0	0	0	2	0
T12	0	0	0	0	0	0	0	0	0	0	0	9



Evaluación de Sugerencias

Matriz de confusión: las tres predicciones.

Entre las 3 Sug.	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12
T1	27	1	0	0	0	0	2	0	4	0	1	6
T2	3	21	1	3	4	0	0	0	3	0	0	2
T3	0	2	29	4	0	1	4	1	0	0	3	1
T4	0	6	0	6	10	0	1	0	8	0	17	3
T5	0	0	0	0	15	0	0	0	0	0	0	0
T6	0	0	0	4	0	4	5	3	0	7	0	0
T7	0	0	0	8	0	5	13	9	2	1	0	1
T8	0	0	0	5	1	19	3	17	1	16	2	0
T9	0	0	0	0	0	0	2	0	12	1	5	7
T10	0	0	0	0	0	1	0	0	0	5	0	0
T11	0	0	0	0	0	0	0	0	0	0	2	0
T12	0	0	0	0	0	0	0	0	0	0	0	10



Evaluación de Sugerencias

- Al incrementar el número de sugerencias a tener en cuenta, aumenta el número de artículos correctamente clasificados.
- La calidad de las sugerencias depende de:
 - ✓ Cantidad y calidad de las palabras del texto seleccionado.
 - ✓ Calidad de la base de conocimiento.
 - ✓ Cantidad de sugerencias a obtener.

Licencia

- Affero General Public License (AGPL).
- Copyleft: exige que todas las modificaciones y extensiones del software sean libres.
- Derivada de GPL de GNU, diseñada específicamente para software que se utiliza a través de una red.
- Obliga a distribuir el software modificado bajo los mismos términos.



Trabajos Futuros

- Ampliar la base de conocimiento recolectando mayor cantidad de producción científica accesible (por ej: ND LTD, publicaciones del DOAJ, etc).
 - ✓ Identificar fuentes de información accesibles a través de protocolos estándares (*OAI-PMH*).
 - ✓ Recolectar datos.
 - ✓ Procesar e incorporar datos a la base de conocimiento.
- Mejorar el procesamiento del texto seleccionado (complemento Firefox) para permitir el uso de cualquier caracter en el texto.
- Analizar la factibilidad de manejar más de un idioma (por ej: español, inglés, portugués).

Preguntas



Agradecimientos

Muchas gracias!!!

Contacto

Universidad Nacional del Sur

Biblioteca Central Prof. Nicolás Matijevic

bc@uns.edu.ar