

Periodismo en Bases de Datos

Búsqueda y Obtención de Información Periodística. Protocolo de Análisis y primeros resultados

Por **Juan Linares, Lluís Codina**

Barcelona, 19 de mayo de 2016
UPF - Seminario DigiDoc



DIGIDOC
Grup de Recerca
en Documentació Digital
i Comunicació Interactiva

Elementos conceptuales

El periodismo en bases de datos:

- Se beneficia de la consulta y explotación de informaciones periodísticas retrospectivas
- Las **hemerotecas digitales** y las **bases de datos** de prensa son su objeto de estudio.

Las bases de datos son el instrumento idóneo en diferentes contextos: mayor cantidad de funciones y utilidades para la búsqueda periodística retrospectiva

- **Requieren:**
 - Por el lado del sistema: funciones de búsqueda y obtención de la información de calidad
 - Por el lado del periodista: un conocimiento adecuado de tales funciones
 - Balance: no obstante, debemos exigir siempre la máxima calidad en el **sistema**
- **Las hemerotecas digitales:** exhiben el grado de **compromiso** de los cibermedios en relación a la calidad de la buscabilidad de su propia producción

Objetivos

1. Identificación y caracterización de indicadores para el análisis evaluativo de la calidad de la representación y recuperación de información periodística
2. Evaluación heurística de la calidad y las posibilidades funcionales de los sistemas analizados para dar soporte a:
 - a. Periodismo en Bases de Datos,
 - b. Investigación en Comunicación Social
 - c. Investigación en Humanidades e Historia del Periodismo, etc.
3. Aportación al análisis de la Buscabilidad de las Informaciones Periodísticas.
4. Propuestas para el diseño de la interacción y de componentes de un sistema de recuperación de información en el ecosistema de la información periodística
5. Desarrollo de patterns

SAAC. Sistema Articulado de Análisis de Cybermedios (Codina et al, 2014)

parámetros + indicadores (1:n)

+

Protocolos de Análisis sobre la interactividad, buscabilidad y visibilidad en cybermedios en Linares, Váñez, Rodríguez Martínez (2016), Linares, Codina, Pedraza (2015) y Linares, Codina y Freixa Font (2015)

Indicador 1.1. Búsqueda simple

<i>Definición</i>	El sistema habilita una caja de consulta donde el usuario puede ejercer la búsqueda introduciendo un término único, compuesto o bien una frase escrita en lenguaje natural para ejercer su necesidad de información.
<i>Pregunta de análisis</i>	¿Dispone el sistema de una caja de consulta para que el usuario pueda ejercer una búsqueda determinada?
<i>Puntuación</i>	0 = No 1 = Sí
<i>Ejemplo</i>	 A screenshot of a Google News search result. The search bar at the top contains the text 'EgyptAir hijack'. Below the search bar, there are navigation options for 'U.K. edition' and 'Classic'. The main content area shows a list of 'Top Stories'. The first story is titled 'EgyptAir hijack: Larnaca negotiations continue' and includes a sub-headline: 'Most of the passengers and crew who were on board a hijacked Egyptian plane diverted to Cyprus have been freed: EgyptAir Flight MS181 was taken over after a passenger claimed to be wearing a suicide explosive belt.' There is a small image of the hijacked plane to the right of the headline. Below the headline, there are several links to related news articles from various sources like BBC News, The Guardian, and CNN.

Figura 1: Indicador “Búsqueda simple”. En el ejemplo se observa el caso de Google News UK

Bases de datos

N.	Nombre	Acceso
1	<i>Google News UK</i>	Abierto
2	<i>Factiva</i>	Restringido
3	<i>MyNews</i>	Restringido
4	<i>Lexis Nexis</i>	Restringido
5	<i>Prensa Histórica</i>	Abierto
6	<i>Periodicals Index Online</i>	Restringido



Protocolo

Un total de **3**
parámetros y **36**
indicadores



Universitat
Pompeu Fabra
Barcelona

DIGIDOC

Grup de Recerca
en Documentació Digital
i Comunicació Interactiva

Parámetros e indicadores

Parámetro 1. "Consulta"

12 indicadores

- Permite expresar sin ambigüedad (en su caso) la necesidad de información del usuario

[Indicador 1.1. Búsqueda simple](#)

[Indicador 1.2. Búsqueda parametrizada](#)

[Indicador 1.2.1. Búsqueda parametrizada por colecciones](#)

[Indicador 1.2.2. Búsqueda parametrizada por formatos](#)

[Indicador 1.2.3. Búsqueda parametrizada por Secciones](#)

[Indicador 1.3. Operadores booleanos](#)

[Indicador 1.4. Índices](#)

[Indicador 1.5. Tendencias de búsqueda](#)

[Indicador 1.6. Autocompletado](#)

[Indicador 1.7. Actualización en tiempo real](#)

[Indicador 1.12. Reutilización de búsquedas](#)

[Indicador 1.12. Ayudas](#)

Parámetros e indicadores

Parámetro 2. “Página de resultados”

20 indicadores

- Permite ordenar, filtrar, exportar... los resultados

[Indicador 2.1. Ordenación](#)

[Indicador 2.2. Filtrado](#)

[Indicador 2.3. Alertas](#)

[Indicador 2.4. Número de resultados por página](#)

[Indicador 2.5. Número total de elementos recuperados](#)

[Indicador 2.6. Agrupación de resultados](#)

[Indicador 2.7. Resumen de búsqueda](#)

[Indicador 2.8. Resultados duplicados](#)

[Indicador 2.9. Exportación selectiva](#)

[Indicador 2.10. Modificar la búsqueda](#)

[Indicador 2.11. Minería de datos 1 - Visualización de la información](#)

[Indicador 2.12. Minería de datos 2 - Autoría](#)

[Indicador 2.13. Minería de datos 3 - Fecha](#)

[Indicador 2.14. Minería de datos 4 - Fuentes](#)

[Indicador 2.15. Minería de datos 5 - Regiones](#)

[Indicador 2.16. Minería de datos 6 - Palabras clave](#)

[Indicador 2.17. Minería de datos 7 - Idioma](#)

[Indicador 2.18. Minería de datos 8 - Temas](#)

[Indicador 2.19. Minería de datos 9 - Industrias](#)

[Indicador 2.20. Feedback de los usuarios](#)

Parámetros e indicadores

Parámetro 3. "Ficha de la noticia"

4 indicadores

- Contiene los datos, metadatos, opciones adicionales de análisis... de la noticia

[Indicador 3.1. Información - Metadatos](#)

[Indicador 3.2. Artículos relacionados \(navegación semántica\)](#)

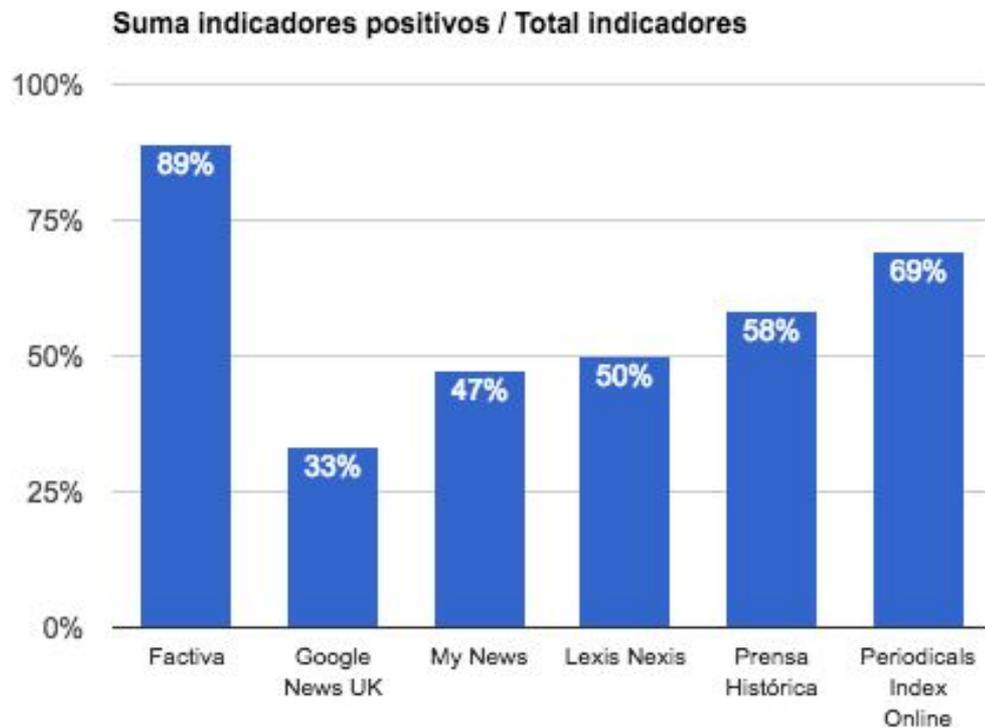
[Indicador 3.3. Identificación de los términos de búsqueda en el documento](#)

[Indicador 3.4. Exportación](#)

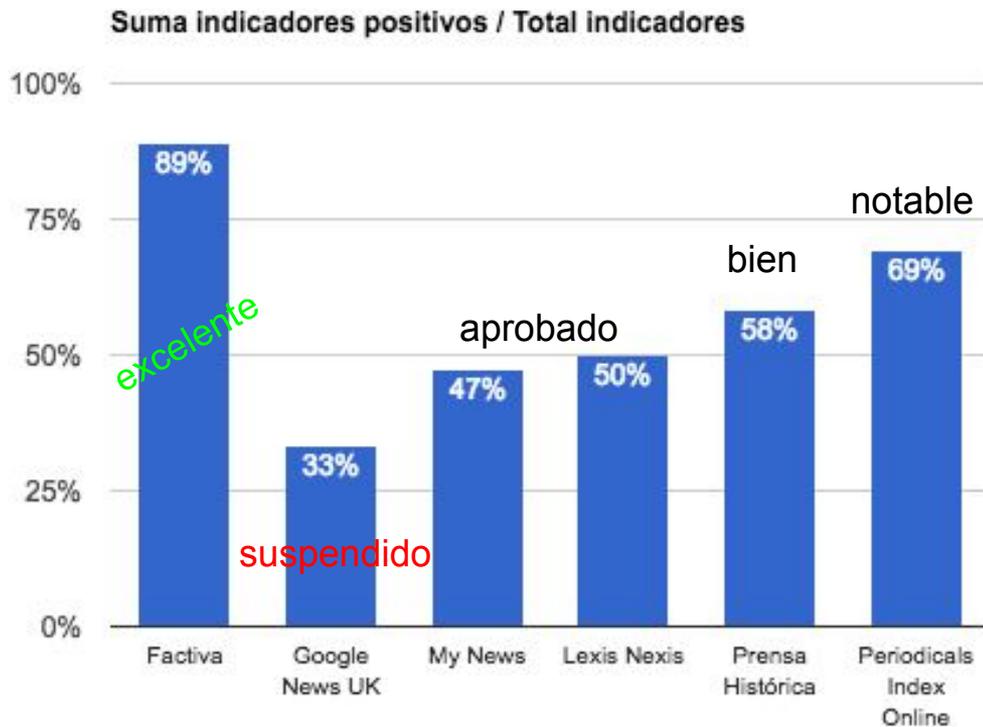
	N.
P1: Consulta	12
P2: Página de resultados	20
P3: Ficha de noticia	4
Total indicadores	36
Total de medios	6

Figura 1. Indicadores y parámetros utilizados

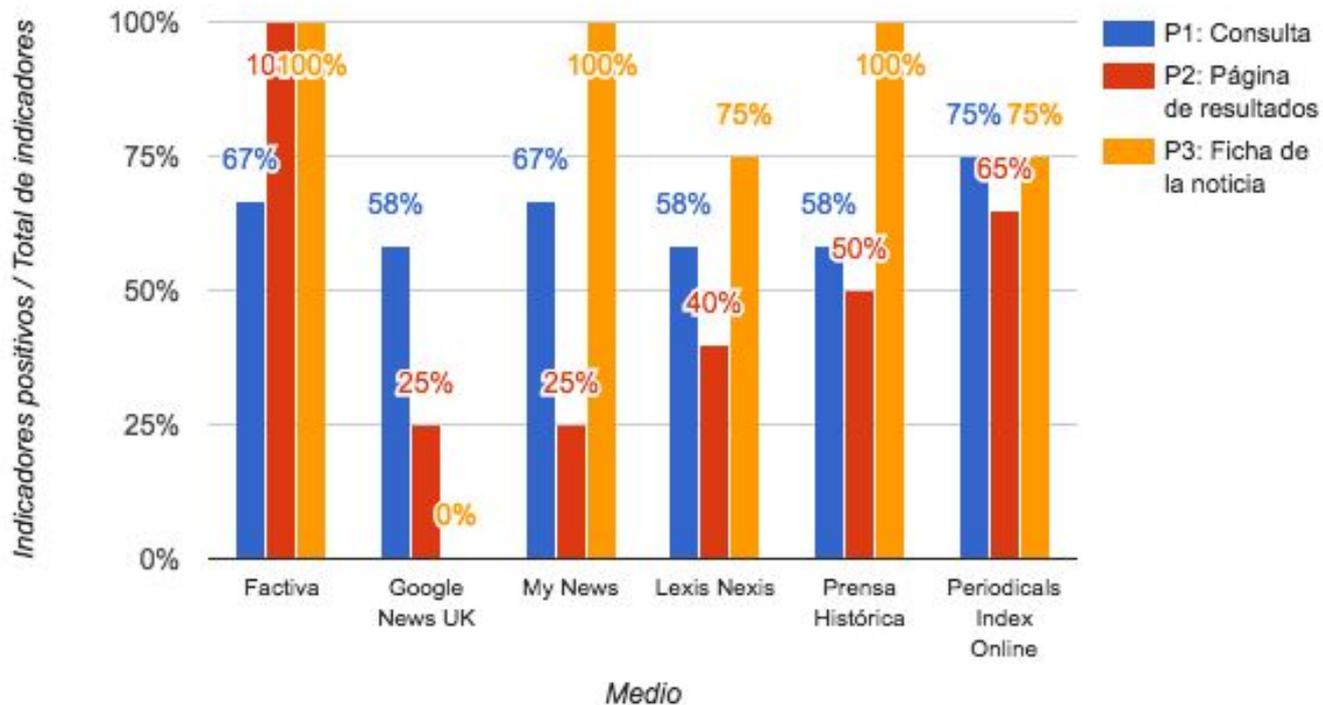
Valoración general. Grado de cumplimiento relativo



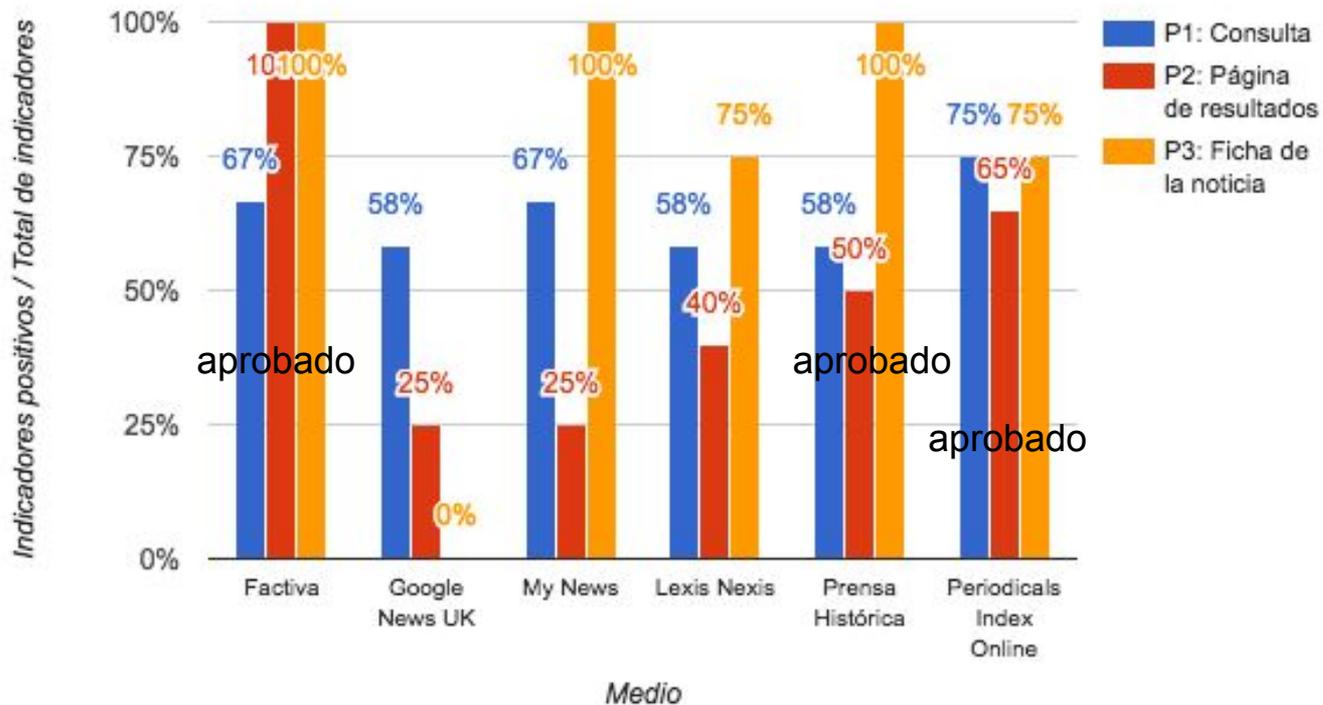
Valoración general



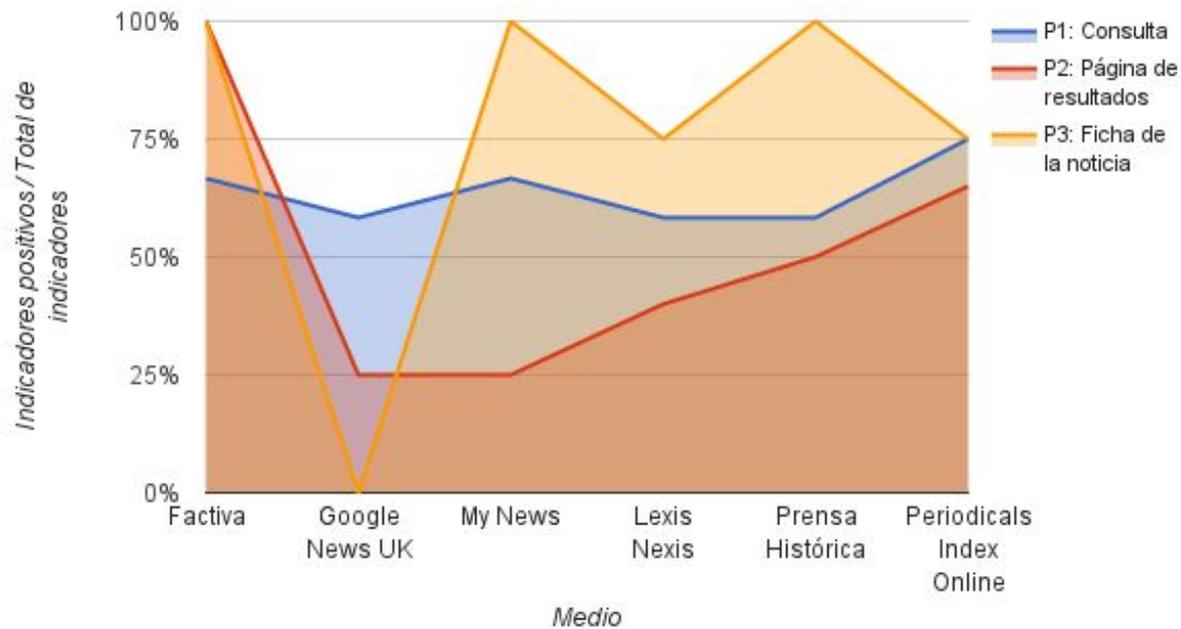
Desglose por parámetros



Desglose por parámetros

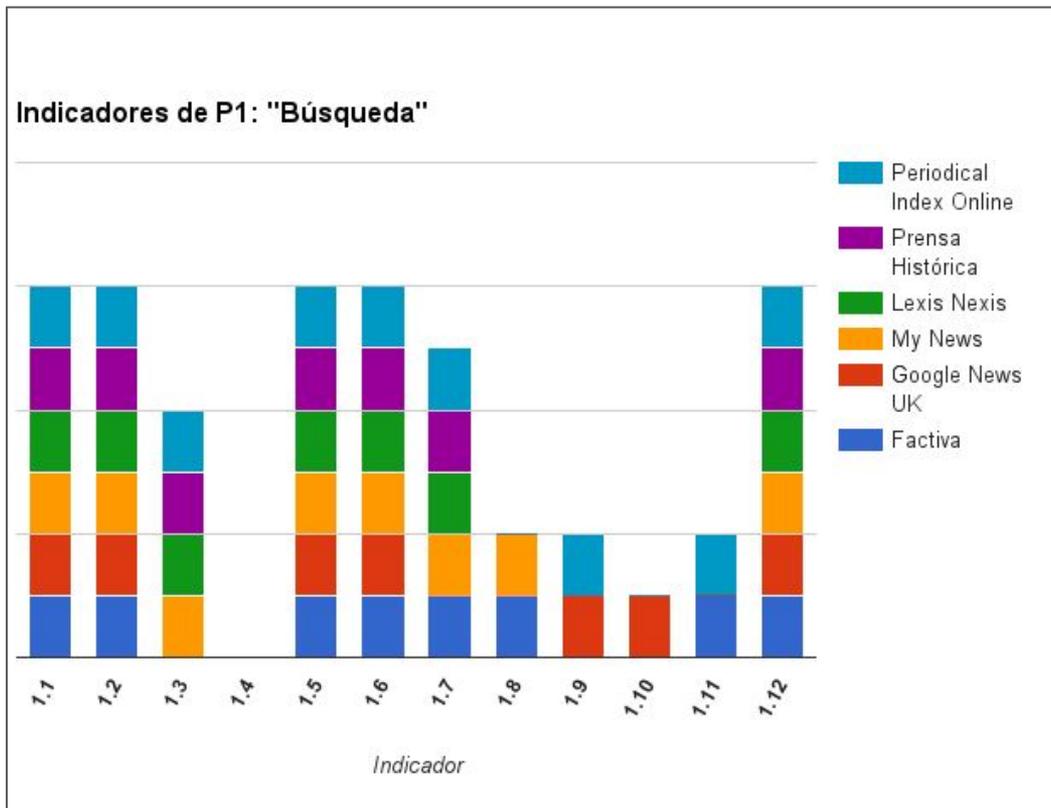


Desglose por parámetros.



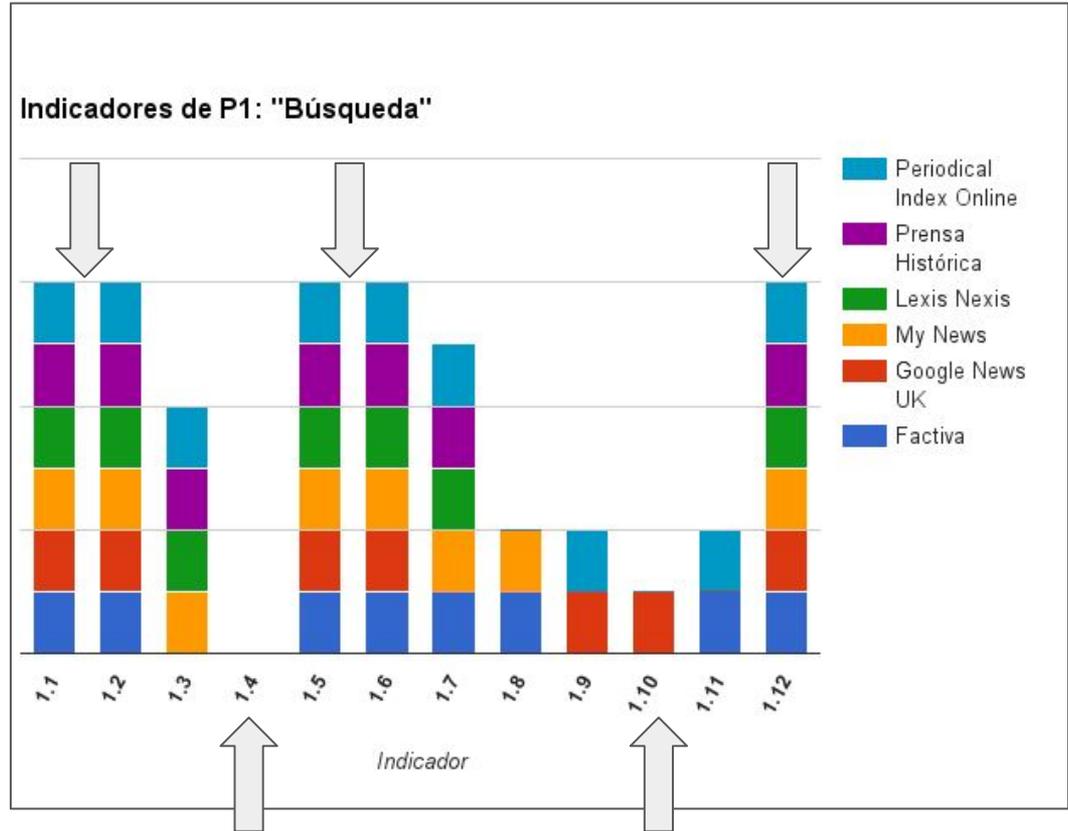
Desglose por indicadores

ID	Indicador
1.1	Búsqueda simple
1.2	Búsqueda parametrizada
1.3	Búsqueda parametrizada por colecciones
1.4	Búsqueda parametrizada por formatos
1.5	Búsqueda parametrizada por secciones
1.6	Operadores booleanos
1.7	Índices
1.8	Tendencia de búsqueda
1.9	Autocompletado
1.10	Actualización en tiempo real
1.11	Reutilización de búsquedas
1.12	Ayudas



Desglose por indicadores

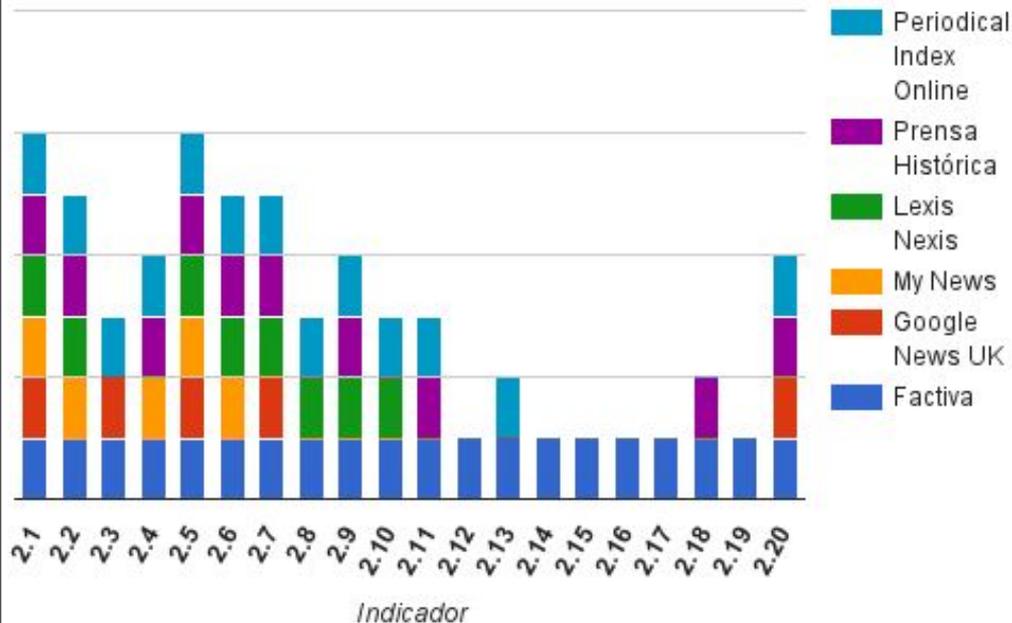
ID	Indicador
1.1	Búsqueda simple
1.2	Búsqueda parametrizada
1.3	Búsqueda parametrizada por colecciones
1.4	Búsqueda parametrizada por formatos
1.5	Búsqueda parametrizada por secciones
1.6	Operadores booleanos
1.7	Índices
1.8	Tendencia de búsqueda
1.9	Autocompletado
1.10	Actualización en tiempo real
1.11	Reutilización de búsquedas
1.12	Ayudas



Desglose por indicadores

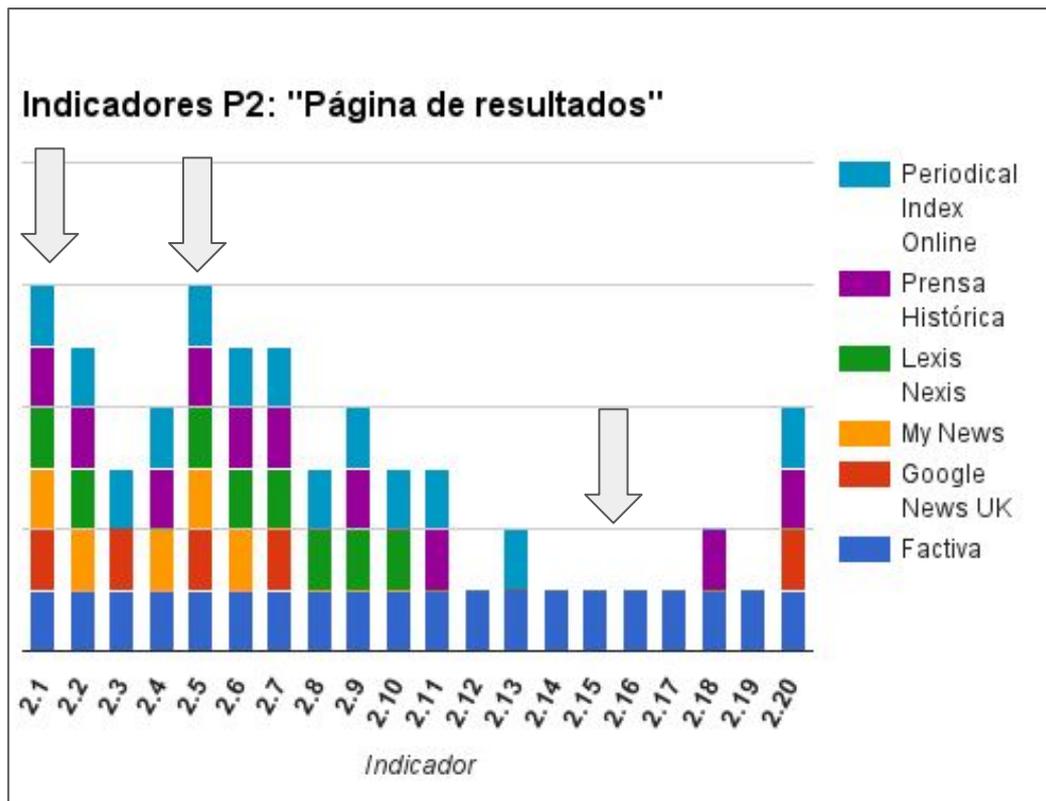
ID	Indicador
1.1	Ordenación
1.2	Filtrado
1.3	Alertas
1.4	Número de resultados por página
1.5	Número total de elementos recuperados
1.6	Agrupación de resultados
1.7	Resumen de búsqueda
1.8	Resultados duplicados
1.9	Exportación selectiva
1.10	Modificar la búsqueda
1.11	Minería de datos 1 - Visualización de la información
1.12	Minería de datos 2 - Autoría
1.13	Minería de datos 3 - Fecha
1.14	Minería de datos 4 - Fuentes
1.15	Minería de datos 5 - Regiones
1.16	Minería de datos 6 - Palabras clave
1.17	Minería de datos 7 - Idioma
1.18	Minería de datos 8 - Temas
1.19	Minería de datos 9 - Industrias
1.20	Feedback de los usuarios

Indicadores P2: "Página de resultados"



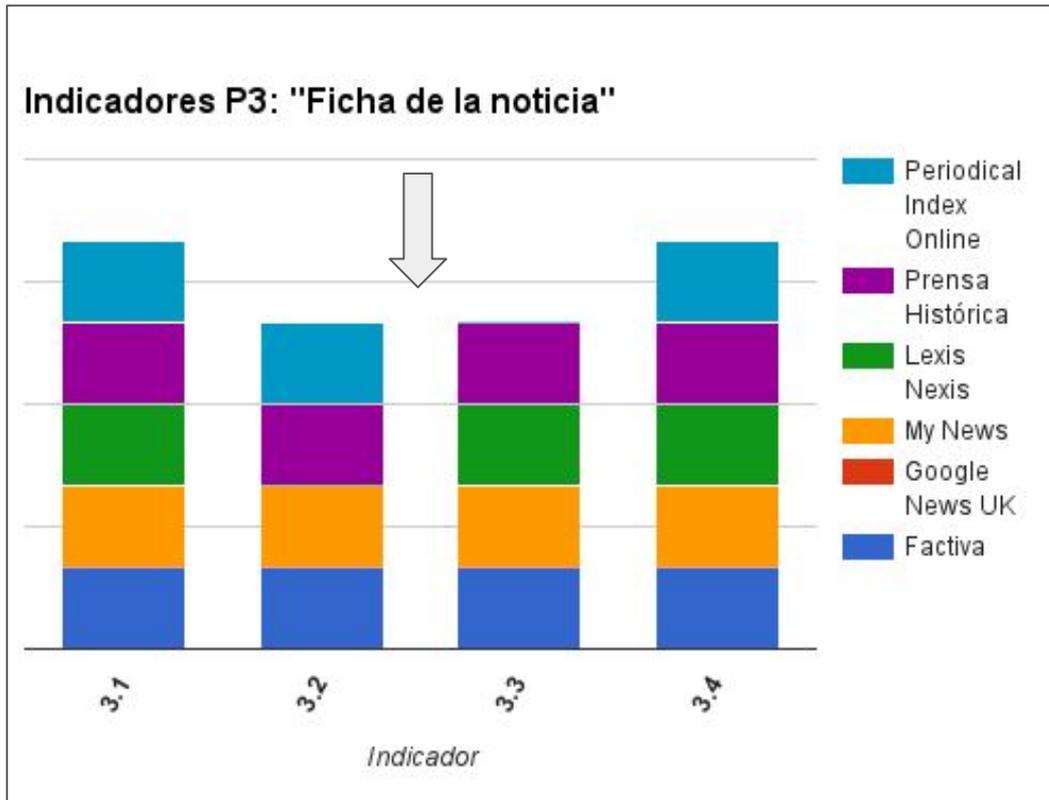
Desglose por indicadores

ID	Indicador
1.1	Ordenación
1.2	Filtrado
1.3	Alertas
1.4	Número de resultados por página
1.5	Número total de elementos recuperados
1.6	Agrupación de resultados
1.7	Resumen de búsqueda
1.8	Resultados duplicados
1.9	Exportación selectiva
1.10	Modificar la búsqueda
1.11	Minería de datos 1 - Visualización de la información
1.12	Minería de datos 2 - Autoría
1.13	Minería de datos 3 - Fecha
1.14	Minería de datos 4 - Fuentes
1.15	Minería de datos 5 - Regiones
1.16	Minería de datos 6 - Palabras clave
1.17	Minería de datos 7 - Idioma
1.18	Minería de datos 8 - Temas
1.19	Minería de datos 9 - Industrias
1.20	Feedback de los usuarios



Desglose por indicadores

ID	Indicador
3.1	Información - Metadatos
3.2	Artículos relacionados (navegación semántica)
3.3	Identificación de los términos de búsqueda en el documento
3.4	Exportación



Siguientes fases

- Análisis:
 - Comparativo aplicado a Hemerotecas Digitales de Cibermedios de referencia
 - Análisis combinado:
 - Funciones (obtenidas de este estudio)
 - Casos
 - Buscar los artículos publicados por un autor
 - Buscar un autor por apellido coincidente con personajes o entidades (p.e. José Luis Barcelona)
 - Buscar por palabras clave que combinen conceptos y sinónimos o variaciones
 - Etc.
- Patterns:
 - Desarrollo de los principales patterns que afectan a la buscabilidad:
 - Búsqueda simple
 - Búsqueda avanzada
 - Página de resultados

Referencias

Juan Linares, Lluís Codina, Pere Freixa. *Tendencias en Cibermedios: Análisis de la Interactividad*. Barcelona: Serie Editorial DigiDoc (UPF), 2015 http://repositori.upf.edu/bitstream/handle/10230/24636/Analisis_ONA-2015.pdf?sequence=1

Juan Linares, Lluís Codina, Rafael Pedraza. *Interactividad en cibermedios: Propuesta nuclear de protocolo de análisis*. Barcelona: Serie Editorial DigiDoc (UPF), 2015. http://repositori.upf.edu/bitstream/handle/10230/23167/protocolo_interactividad_2015.pdf?sequence=1

Lluís Codina (2016). *Periodismo en Bases de Datos: Características y Funciones de Búsqueda. El caso de Factiva y de LexisNexis* <http://www.lluiscodina.com/periodismo-bases-datos-factiva-lexisnexis/>

Lluís Codina; Rafael Pedraza; Javier Díaz Noci; Ruth Rodríguez-Martínez; Mario Pérez-Montoro; Víctor Cavaller-Reyes. "Sistema Articulado de Análisis de Cibermedios (SAAC): Una propuesta sobre el qué y el cómo para estudiar medios de comunicación digitales." *Hipertext.net* [online], 2014, Núm. 12. <http://raco.cat/index.php/Hipertext/article/view/275560/364530> DOI: 10.2436/20.8050.01.13

Linares, Juan; Codina, Lluís; Vàllez, Mari y Rodríguez-Martínez, Ruth. *Interactividad, Buscabilidad y Visibilidad en Cibermedios: Sistema de Análisis y Resultados*. Barcelona: Departamento de Comunicación. Serie Editorial DigiDoc, 2016 http://repositori.upf.edu/bitstream/handle/10230/26280/Linares_inter_2016.pdf



Universitat
Pompeu Fabra
Barcelona

DIGIDOC
Grup de Recerca
en Documentació Digital
i Comunicació Interactiva