



Universitat
Pompeu Fabra
Barcelona

DIGIDOC
Grup de Recerca
en Documentació Digital
i Comunicació Interactiva

Periodismo en Base de Datos

Buscabilidad de la
Información Periodística Retrospectiva
Protocolo de Análisis y Resultados

Juan Linares · 14 junio 2016 · Seminario SEO & COM

Elementos conceptuales

Periodismo en Base de datos es aquel que se beneficia de la consulta y explotación de informaciones periódicas publicadas con anterioridad sobre una historia o un ámbito temático determinado (Codina, 2016)

Periodismo en Base de Datos: **funciones**

A través de la información retrospectiva, un periodista puede saber, sobre una determinada historia o hecho noticiable:

- **qué** se ha publicado anteriormente
- **cómo** ha sido tratada la noticia o la historia
- **dónde** se ha publicado anteriormente noticias relacionadas
- **cuáles** son los principales actores, entidades y fuentes que intervienen
- **obtener** nuevas ideas e enfoques
- llevar a cabo una **minería de datos** con capacidad de descubrimiento de nuevas dimensiones

Periodismo en Base de Datos: **dimensiones**

- **Competencias informacionales.** Lado del periodista: el profesional debe ser capaz de transformar sus necesidades de información en estrategias de búsqueda efectivas.
- **Buscabilidad de la información periodística.** Lado del sistema: aspectos funcionales que permiten la búsqueda y descubrimiento de la información

Foco y Racionalidad de esta investigación

- **Foco:** los aspectos de calidad que están en el lado del *sistema*. La infraestructura sin la cual el Periodismo en Base de Datos no puede existir.
- **Racionalidad:** la calidad de los sistemas de búsqueda como parte del *compromiso* de los actores involucrados: cibermedios y sistemas de información periodística

¿Cómo **analizar** y determinar la calidad en la **buscabilidad** de la información periodística retrospectiva?

Objetivo

O1. Identificar y caracterizar “indicadores” para el análisis evaluativo de la calidad de la buscabilidad y obtención de información en Bases de Datos Periodísticas de referencia

- Diseño, desarrollo y aplicación de un **Protocolo de Análisis de Bases de Datos Periodísticas** de referencia

Muestra

6 Bases de Datos Periódicas de Referencia

- *Google News UK*
- *Factiva*
- *Periodicals Index Online*
- *Lexis Nexis*
- *MyNews*
- *Prensa Histórica*

Muestra

6 Bases de Datos Periódicas de Referencia

- *Google News UK*
 - Activa desde el 2002 (beta), 2006 (lanzamiento)
 - Acceso abierto
 - 4500 fuentes periódicas (en inglés)
 - Información retrospectiva: 30 días antes
 - Google



Muestra

6 Bases de Datos Periódicas de Referencia

- *Factiva*
 - Activa desde 1999
 - Acceso restringido
 - 32.000 fuentes periódicas
 - Down Jones

The logo for Factiva, featuring the word "FACTIVA" in a bold, blue, sans-serif font. A registered trademark symbol (®) is located at the top right of the letter "A".

Muestra

6 Bases de Datos Periódicas de Referencia

- *Lexis Nexis*
 - Activa desde 1977
 - Acceso restringido
 - 45.000 fuentes periódicas
 - REED Elsevier



Muestra

6 Bases de Datos Periódicas de Referencia

- *My News*
 - Activa desde 1995
 - Acceso restringido
 - Centenar(es) fuentes periódicas (España)
 - 190 millones de artículos indizados



Muestra

6 Bases de Datos Periódicas de Referencia

- *Periodicals Index Online*
 - Activa desde 1995?
 - Acceso restringido
 - 6000 revistas (Humanidades y Ciencias Sociales)
 - 20 millones de artículos (desde 1665)
 - ProQuest

Periodicals Index Online

Muestra

6 Bases de Datos Periódicas de Referencia

- *Prensa Histórica*
 - Activa desde 20xx?
 - Acceso abierto
 - +1 millón de ejemplares digitalizados
 - Primer ejemplar: 1753
 - Ministerio de Educación Cultura y Deporte (ESP)



Protocolo de Análisis: referencias

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

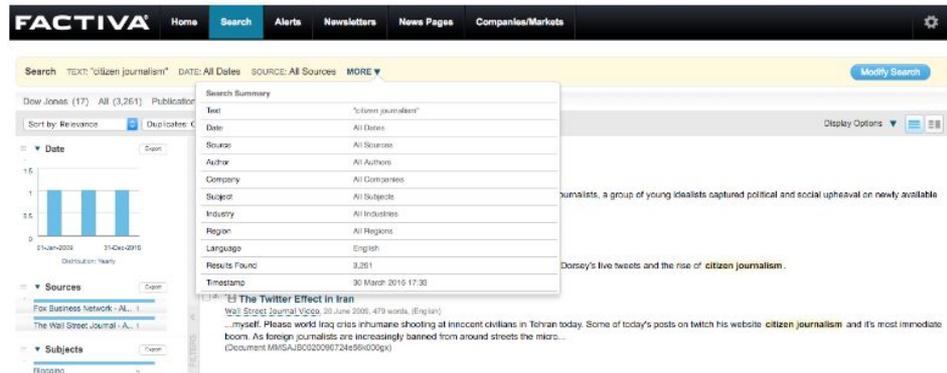
Entregables Serie Editorial DigiDoc: Protocolos de Análisis

Linares, Vállez, Rodríguez Martínez (2016), Linares, Codina, Pedraza (2015) y Linares, Codina y Freixa Font (2015)

Fundamentos del SAAC

- El análisis se articula en parámetros + indicadores
 - Parámetro = “grano grueso”. Qué queremos estudiar (p.e: “consulta”)
 - Indicadores = “grano fino”. Cómo vamos a realizar el estudio. Caracterizados en Fichas Sistemáticas (p.e: “búsqueda avanzada”) -> evitar ambigüedades
- Relación P : i = 1:n

Indicador 2.7. Resumen de búsqueda

<i>Definición</i>	El sistema permite visualizar los parámetros que se han utilizado en la búsqueda.
<i>Pregunta de análisis</i>	¿Se pueden consultar todos los parámetros utilizados en una búsqueda?
<i>Puntuación</i>	0 = No 1 = Sí
<i>Ejemplo</i>	<p><i>Factiva</i> dispone de la opción “MORE” que permite visualizar todos los parámetros utilizados en la búsqueda.</p>  <p>The screenshot shows the Factiva search interface. At the top, there is a navigation bar with 'FACTIVA' and links for Home, Search, Alerts, Newsletters, News Pages, and Companies/Markets. Below this is a search bar with the text 'citizen journalism' and a 'Modify Search' button. A 'Search Summary' dropdown menu is open, listing various filters: Text, Date, Source, Author, Company, Subject, Industry, Region, Language, Results Found (8,261), and Timestamp (30 March 2016 17:30). The main search results area shows a bar chart for 'Date' and a list of sources including 'Fox Business Network', 'The Wall Street Journal', and 'Dorsey's live tweets and the rise of citizen journalism'.</p>

3 parámetros

36 indicadores

P1. Consulta

P2. Página de Resultados

P3. Ficha de la noticia

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](#)



Search TEXT: 15m or "15-m" or spanishrev... DATE: 01/02/2011 to 01/11/2011 SOURCE: El Pais - All sources Or El M... MORE ▾

Cancel

Free Text Search

Query Genius ⓘ ▾

Search Form

15m or "15-m" or spanishrevolution or "no les votes" or "nolesvotes" or "juventud sin futuro" or "democracia real ya" or "Plataforma de coordinación de grupos pro-movilización ciudadana" or "acampada" or "acampadas"

Examples

215/2048

 Include additional Blogs and BoardsDate ▾ 01 / 02 / 2011 01 / 11 / 2011 Duplicates ▾

Search

Source El Pais - All sources ▾ Or El Mundo - All sources ▾ Or ABC (Spain) - All sources ▾ Or La Vanguardia (Catalonia, Spain, Spanish Language) ▾ Or El Periodico - All sources ▾ Or 20 Minutos (Spain, Spanish Language) ▾Author All AuthorsCompany All CompaniesFactiva Expert Search Subject All SubjectsIndustry All IndustriesRegion All RegionsLook up Language Spanish

More Options ▾

Search for free-text terms in:

 ▾

Exclude:

- Republished news ⓘ
- Recurring pricing and market data ⓘ
- Obituaries, sports, calendars... ⓘ

Sort results by:

 ▾

Language ▶ Catalan ✕ Spanish ✕

More Options ▼

Search for free-text terms in:

- Full Article
- Headline and Lead Paragraph
- Headline**
- Byline

Exclude:

- Recurring pricing and market data ⓘ
- Obituaries, sports, calendars... ⓘ

Sort results by: Relevance ⌵

Start New Search Search

[Dow Jones](#) (3) [All](#) (20,479) [Publications](#) [Web News](#) (0) [Blogs](#) (0) [Pictures](#) (0) [Multimedia](#) (0)

Indicador 1.2.3. Búsqueda parametrizada por Secciones. Fuente: Factiva

Font de Noticia

Posició: Tot l'article

Secció: TOTES

Edició: CULTURA

Autor: INTERNACIONAL

Mitjà/Diari: NACIONAL

Rellevància: PORTADA

més percentatge més exigència en la coincidència dels resultats)

Cercar

Indicador 1.2.3. Búsqueda parametrizada por Colecciones. Fuente: MyNews

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](#)

Search TEXT: 15m or "15-m" or spanishrev... DATE: 01/02/2011 to 01/11/2011 SOURCE: El Pais - All sources Or El M... MORE ▾

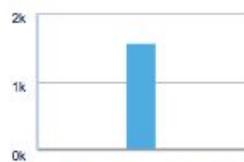
Modify Search

Dow Jones (0) All (1,539) **Publications** Web News (0) Blogs (0) Pictures (0) Multimedia (0)

Sort by: Oldest first Duplicates: Identical

Display Options ▾

▼ **Date**



Distribution: Yearly

▼ **Companies**

Parliament of Catalonia	31
Catalonia Department of t...	7
Centre of Sociological Re...	3
Unión General de Trabaja...	3
Universidad de Sevilla	3
Andalusia Ministry of Envi...	2
Caja de Ahorros de Valen...	2
Caja de Ahorros y Monte ...	2
Nueva Rumasa S.A.	2

▼ **Sources**

El Mundo - All sources	604
El Periodico - All sources	390
El Pais - All sources	356

Headlines 1 - 100 of 1,539 Next 100 ▶ Total duplicates: 15

1. **No les votes**

[Elmundo.es](#), 9 February 2011, 31 words, Carlos S. Almeida, (Spanish)

Si las reglas no se adaptan a las necesidades de la metrópoli, se cambian. Todo seguirá igual mientras no asumamos nuestra culpa. Jaque perpetuo (Document MNDES00020110210e729000by)

2. **#NoLesVotes, más allá de la 'Ley Sinde'**

[Elmundo.es](#), 25 February 2011, 34 words, Sergio Rodríguez, (Spanish)

Lo que comenzó como una protesta contra la norma para cerrar webs se ha convertido en una forma de canalizar protestas ciudadanas. El Catalejo (Document MNDES00020110225e72p0005f)

3. **Dos 'shocks' en una semana y vienen los Idus de marzo; El récord en la subida de los carburantes mete a la economía española en la...dos 'sh...15-M**

[El Mundo](#), 6 March 2011, 1543 words, CARLOS SEGOVIA, (Spanish)

La estanción por la subida del precio de los carburantes y la subida del Euribor dificultan la salida de la crisis económica en España. Cajas de Ahorros. Rouco Varela. Los bonus de Caja Madrid (Document MUNDO00020110306e7360001u)

1 duplicate article(s) identified

4. **Los trabajadores de Sintel celebran el décimo aniversario de su acampada en Castellana**

[Elmundo.es](#), 20 March 2011, 233 words, EP, (Spanish)

La Asociación de Trabajadores de Sintel de León ha organizado la semana conmemorativa del décimo aniversario de la 'Acampada en la Castellana', que acogerá la ciudad entre los días 21 y 26 de marzo.El alcalde de León, Francisco ... (Document MNDES00020110320e73k0009e)

5. **#nolesvotes pasa de protestar por la 'Ley Sinde' a denunciar la corrupción política**

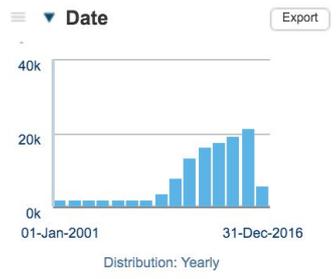
[Elmundo.es](#), 31 March 2011, 692 words, P. Romero, (Spanish)

Todo comenzó como una protesta contra las medidas 'antipiratería' de la Ley de Economía Sostenible impulsadas por el Gobierno. Ahora, el movimiento ciudadano #nolesvotes 'salta' a abordar otros temas más generales, como ... (Document MNDES00020110331e73v000eg)

Search TEXT: "social media" DATE: All Dates SOURCE: All Sources **MORE** ▼ Modify Search

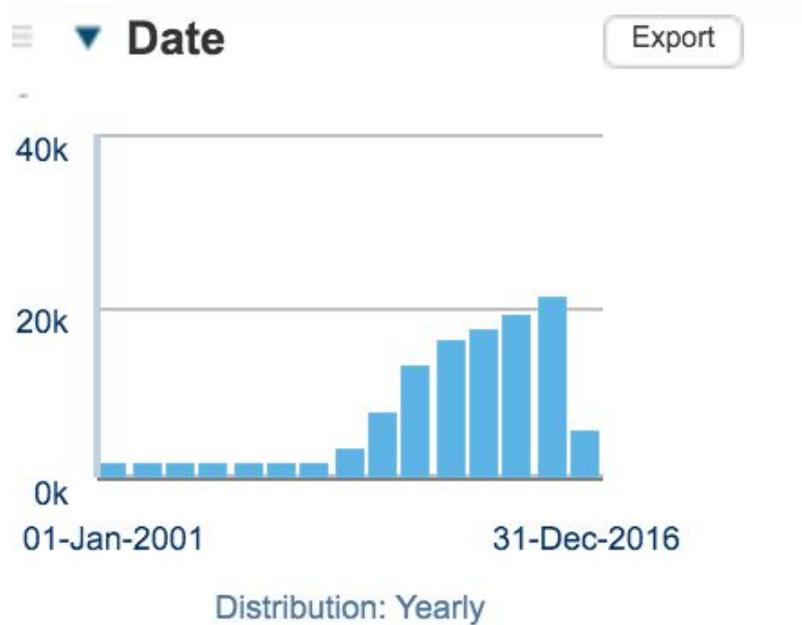
Dow Jones (2,987) All (116,413) **Publications** Web News (12,780) Blogs (1,183) Pictures (143) Multimedia (286)

Sort by: Relevance ▾ Duplicates: Off ▾ 👁️ 📧 🖨️ 📁 RTF 📄 ? Display Options ▼ ☰ ☰



- Headlines 1 - 100 of 102,022 Next 100 ▶
- Social media protections for workers advance in Nebraska**
 Associated Press Newswires, 1:50, 31 March 2016, 142 words, (English)
 LINCOLN, Neb. (AP) — Nebraska lawmakers have advanced new safeguards to keep employers from demanding access to their workers' social media accounts.
 (Document APRS000020160330ec3u00ajl)
 - Social media frenzy over helter-skelter thrills at Trinity Leeds**
 Yorkshire Evening Post, 30 March 2016, 376 words, (English)
 Shoppers at Trinity Leeds will have no excuse for getting bored in town for the next few weeks. A bright yellow helter-skelter slide allows visitors

Indicador 2.10. Exportación selectiva. Fuente: Factiva



☰ ▼ **Companies** Export

Parliament of Catalonia	31
Catalonia Department of t...	7
Centre of Sociological Re...	3
Unión General de Trabaja...	3
Universidad de Sevilla	3
Andalusia Ministry of Envl...	2
Caja de Ahorros de Valen...	2
Caja de Ahorros y Monte ...	2
Nueva Rumasa S.A.	2

☰ ▼ **Sources** Export

El Mundo - All sources	604
El Periodico - All sources	390
El Pais - All sources	356

Indicador 2.10. Minería de Datos 1 - Visualización de la información. Fuente: Factiva

Parámetros e indicadores

Parámetro 3.

“Ficha 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](#)

4 indicadores

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

EL MUNDO.es

No les votes

Carlos S. Almeida

31 words

9 February 2011

Elmundo.es

MNDES

Spanish

(c) Mundinteractivos, S.A.

Si las reglas no se adaptan a las necesidades de la metrópoli, se cambian. Todo seguirá igual mientras no asumamos nuestra culpa.

Jaque perpetuo

Mundinteractivos SA

Document MNDES00020110210e729000by

Nick De Marco and Chris Walsh

526 words

31 March 2016

02:01

thetimes.co.uk

TIMEUK

English

© Times Newspapers Limited 2016

In the digital age, news is fast becoming a free-for-all commodity. Or is it? Phones are the latest device for uploading news

Fanatix, a sports clips company, launched an app and website allowing fans to upload and share, on a near-live basis, short clips of sports broadcast footage. Users could search for their favourite sport or team and view the latest action for free.

...

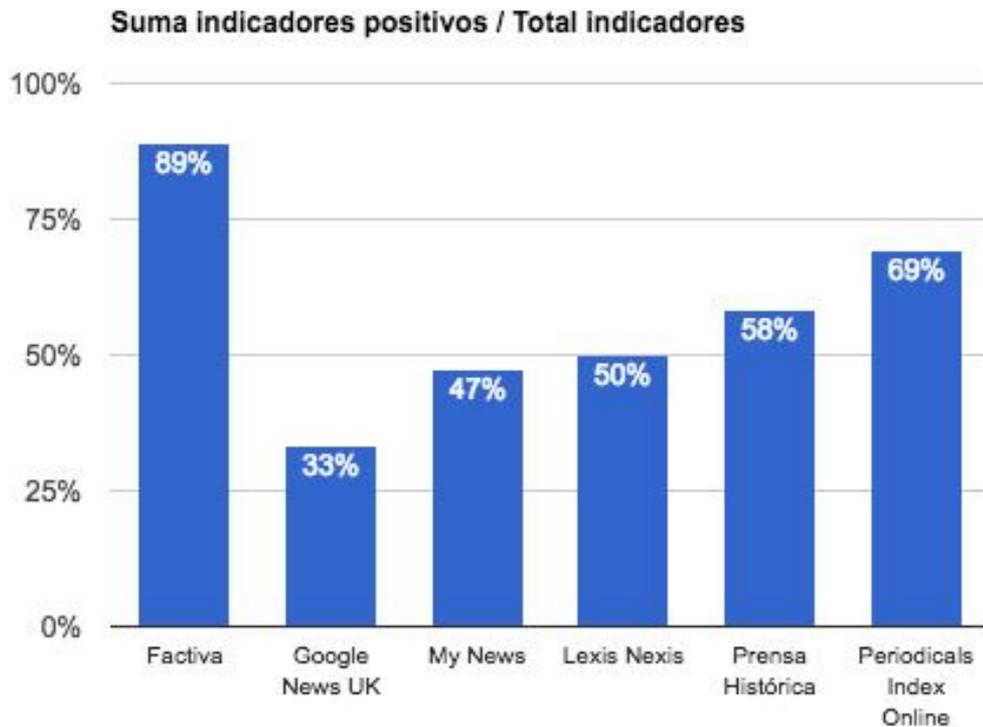
They argued that users could only post eight-second clips, with commentary and attribution, viewable for 24 hours, and that in the “citizen **journalism**” era where people film, share and view news events through their mobiles and online, the defence ought to be as open to them as it is to broadcasters...

...

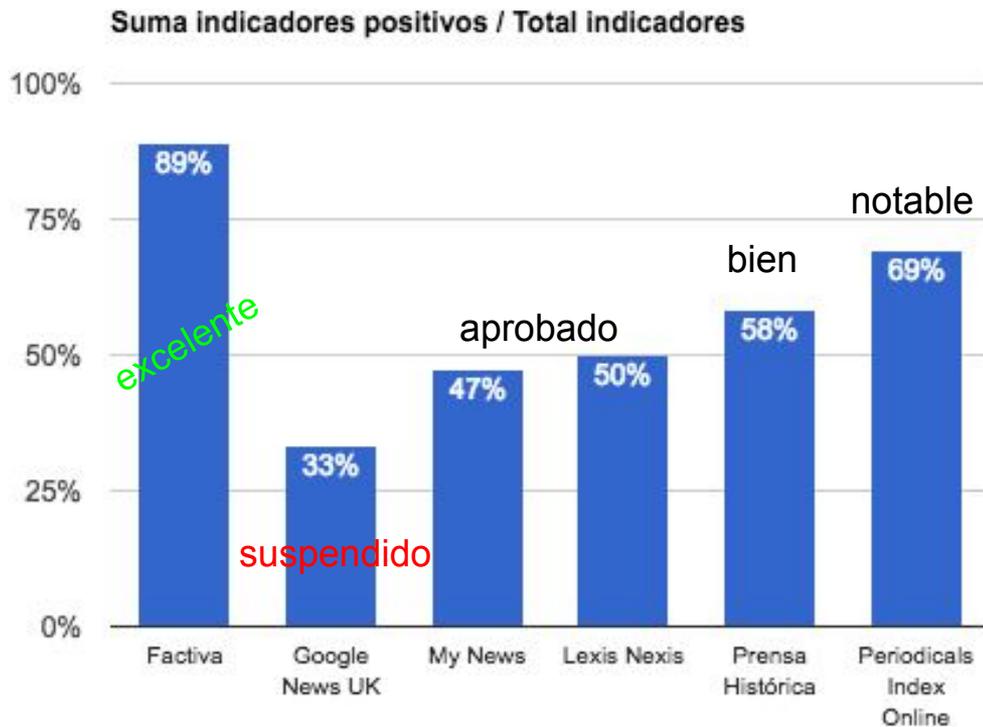
Indicador 3.3. Identificación de los términos de búsqueda en el documento

Resultados

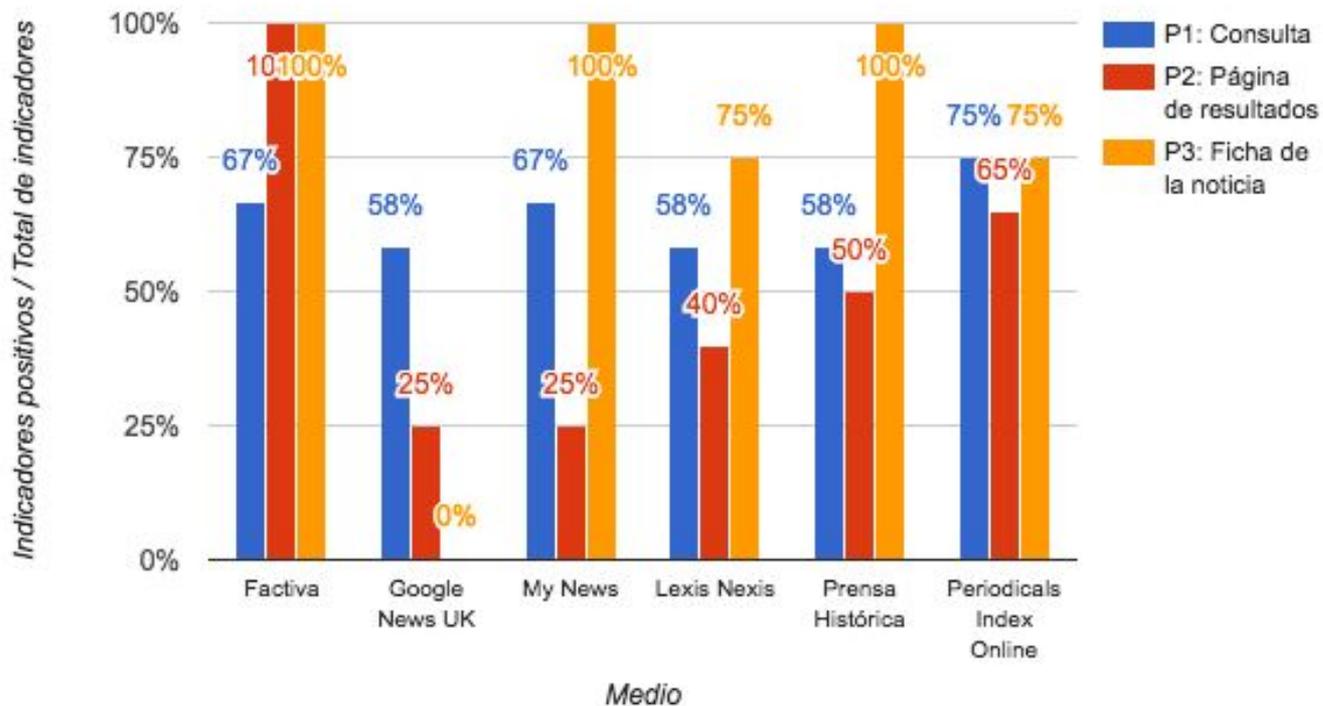
Valoración general. Grado de cumplimiento relativo



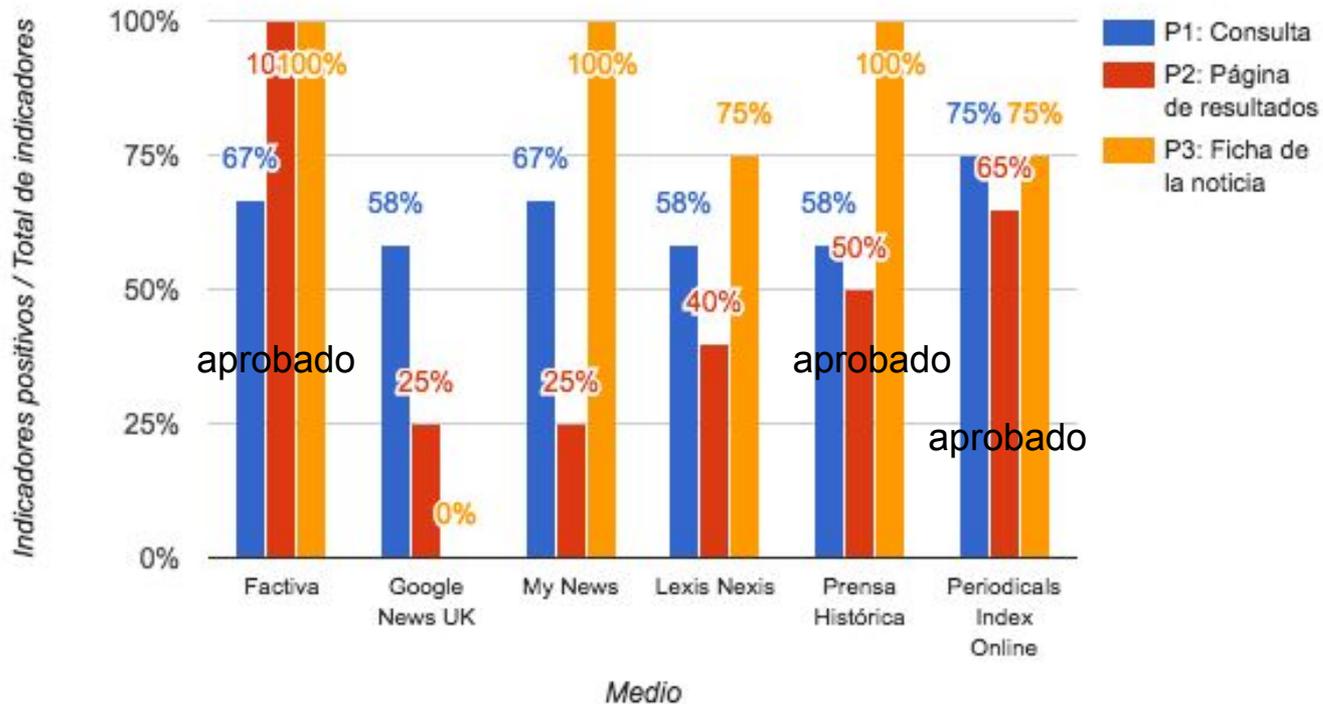
Valoración general. Grado de cumplimiento relativo



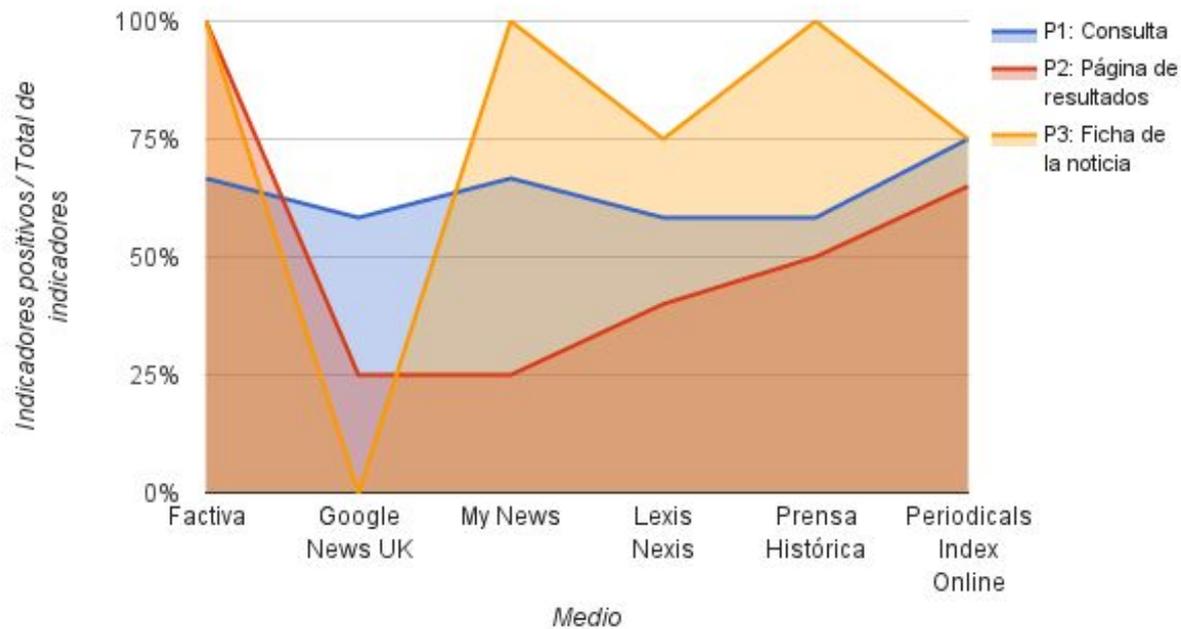
Desglose por parámetros



Desglose por parámetros

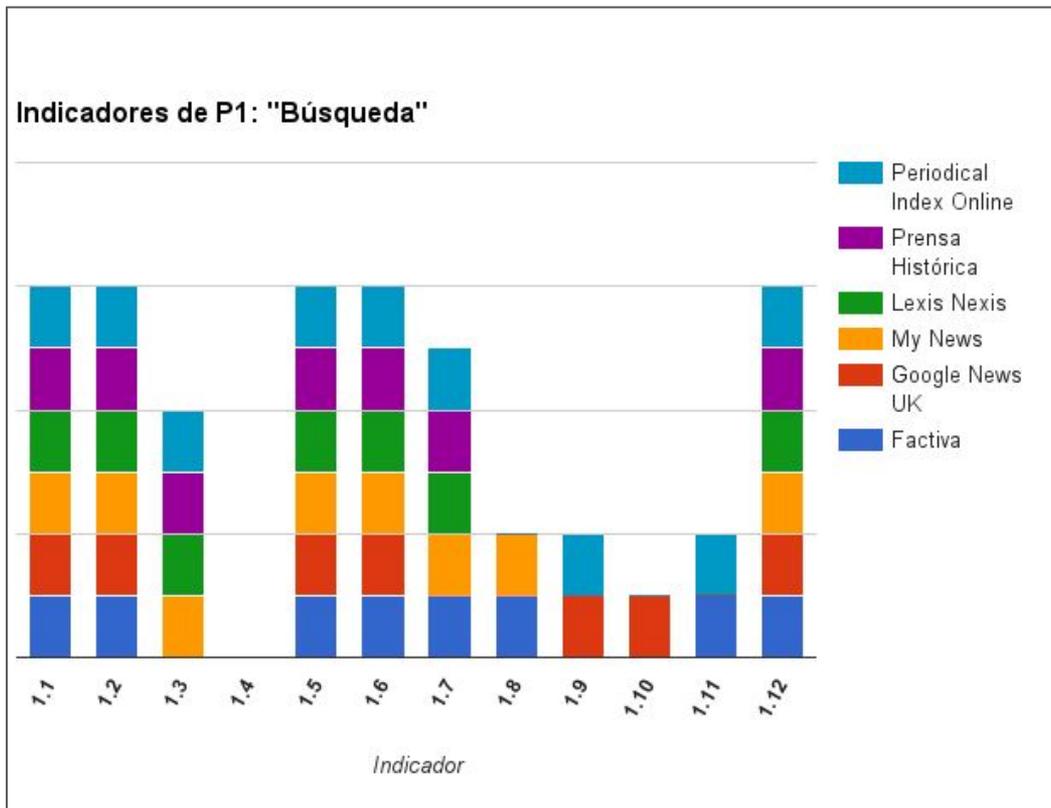


Desglose por parámetros.



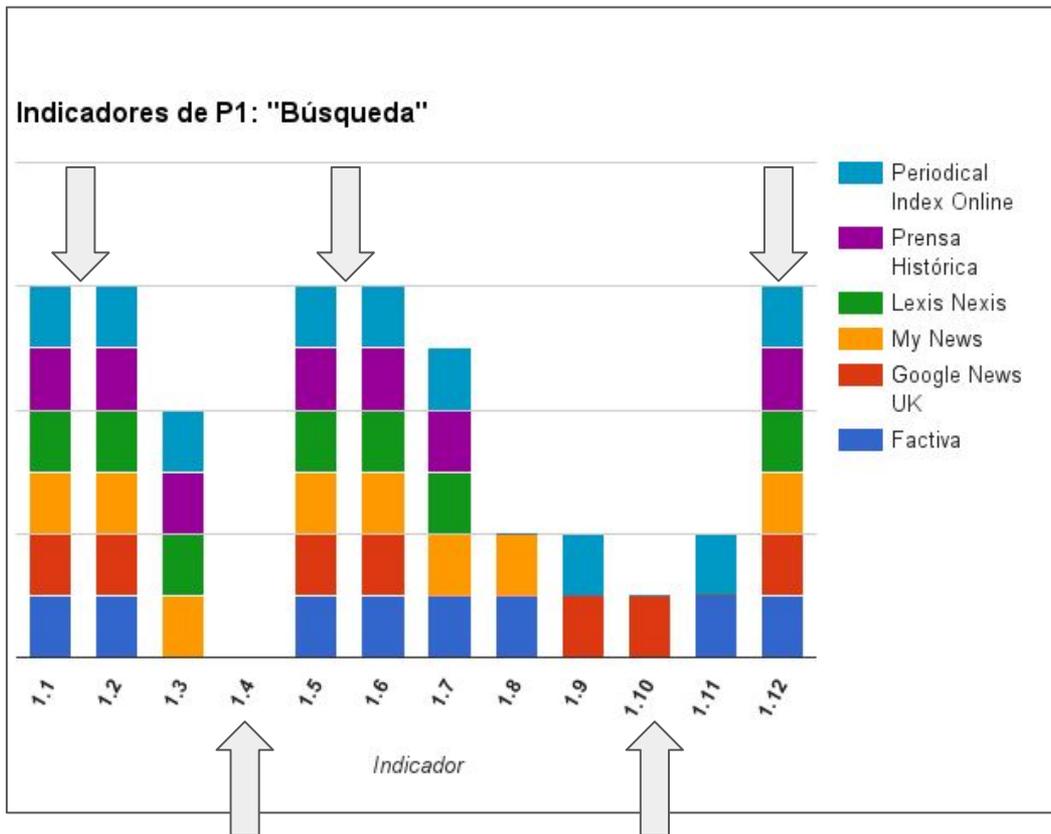
Desglose por indicadores

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

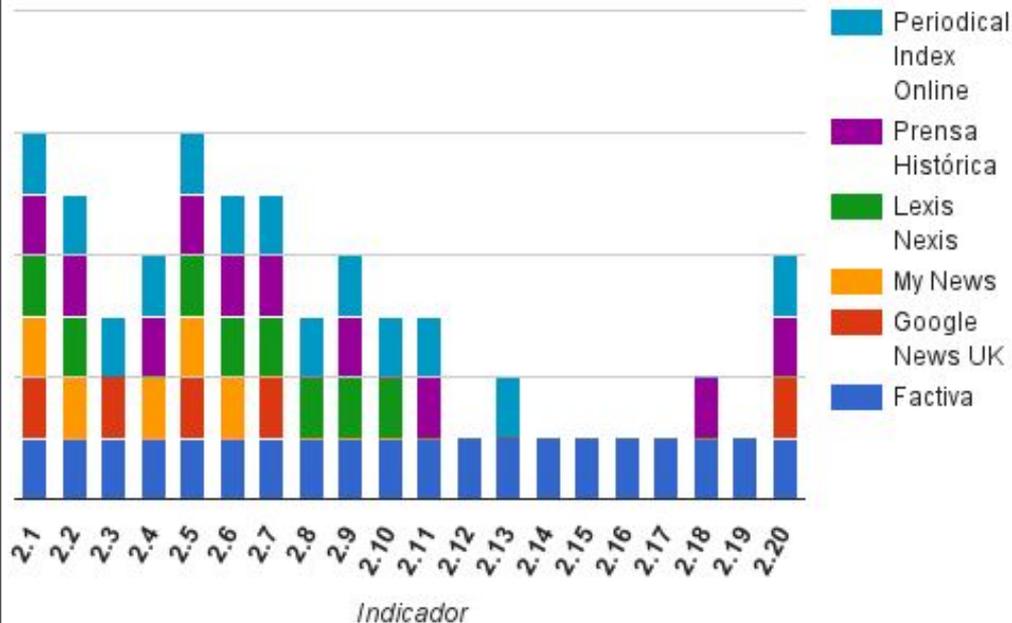
ID	Indicador
1.1	Búsqueda simple
1.2	Búsqueda parametrizada
1.3	Búsqueda parametrizada por colecciones
1.4	Búsqueda parametrizada por formatos
1.5	Búsqueda parametrizada por secciones
1.6	Operadores booleanos
1.7	Índices
1.8	Tendencia de búsqueda
1.9	Autocompletado
1.10	Actualización en tiempo real
1.11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



Desglose por indicadores

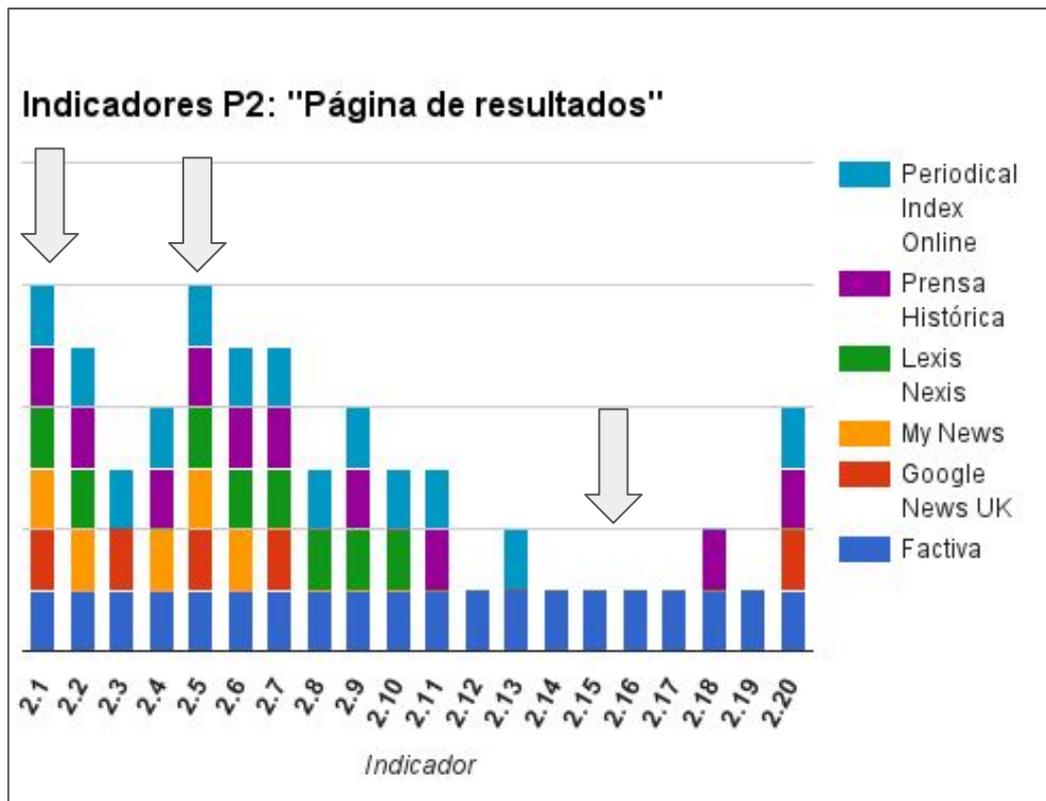
ID	Indicador
1.1	Ordenación
1.2	Filtrado
1.3	Alertas
1.4	Número de resultados por página
1.5	Número total de elementos recuperados
1.6	Agrupación de resultados
1.7	Resumen de búsqueda
1.8	Resultados duplicados
1.9	Exportación selectiva
1.10	Modificar la búsqueda
1.11	Minería de datos 1 - Visualización 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"



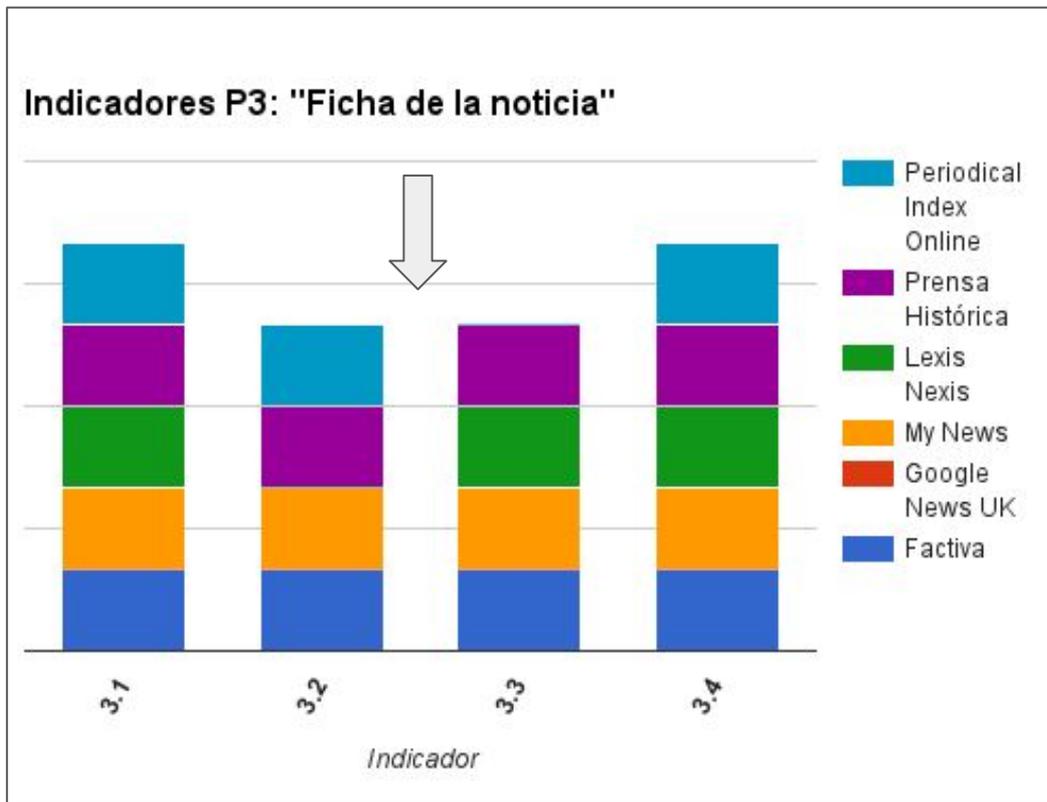
Desglose por indicadores

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



Conclusiones

Conclusiones

- Escalabilidad del SAAC: de Cibermedios a Bases de Datos
- Bases de Datos Periódicas son instrumento idóneo para la búsqueda periodística retrospectiva que requieren un equilibrio
 - Por el lado del sistema: funciones de búsqueda y obtención de resultados de calidad
 - Por el lado del periodista/investigador: conocimiento adaptado a estas funciones
- En cuanto al grado de cumplimiento relativo de indicadores, dos modelos “canónicos”: *Factiva* y *Periodicals Index Online*

Conclusiones

- Las Bases de Datos Periódicas analizadas se presentan como sistemas...:
 - Principalmente orientados a satisfacer las funciones de búsqueda en la página de inicio (consulta)
 - Que presentan, comparativamente, una alta variabilidad en cuanto a la calidad en las opciones de recuperación de resultados
 - Destaca “Minería de datos” como indicador relevante
 - Que podrían mejorar sustancialmente la calidad en cuanto al número de indicadores en sus registros
 - Destaca que Google News no presenta “fichas de la noticia”

Juan Linares, Lluís Codina, Pere Freixa: "Tendencias en Cibermedios. Análisis de la Interactividad". Barcelona: *Serie Editorial DigiDoc*, 2015 <http://hdl.handle.net/10230/24636>

Juan Linares, Lluís Codina, Rafael Pedraza: "Interactividad en Cibermedios: Propuesta nuclear de protocolo de análisis". Barcelona: *Serie Editorial DigiDoc*, 2015 <https://repositori.upf.edu/handle/10230/23167>

Guallar, Javier; Abadal, Ernest; Codina, Lluís (2012). "Sistema de análisis de hemerotecas de prensa digital". *Trípodos*, n. 31, pp. 37-64. <http://eprints.rclis.org/19929/>

Guallar, Javier; Abadal, Ernest; Codina, Lluís (2012). "Hemerotecas de prensa digital: evolución y tendencias". *El profesional de la información*, vol. 21, n. 6, pp. 595-605. <http://eprints.rclis.org/18199/>

Lluís Codina (2016). *Periodismo en Bases de Datos: Características y Funciones de Búsqueda. El caso de Factiva y de LexisNexis* [Online]. <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*, 2014, 12
<http://raco.cat/index.php/Hipertext/article/view/275560/364530>

Marti. A. Hearst (2009). *Search User Interfaces*. Cambridge [etc.]: Cambridge University Press.

Lluís Codina (14 marzo 2016). Periodismo en Bases de Datos: Características y Funciones de Búsqueda. El caso de Factiva y de LexisNexis. lluis codina upf · docencia e investigación en comunicación y documentación digital <http://www.lluiscodina.com/periodismo-bases-datos-factiva-lexisnexis/>
<http://www.lluiscodina.com/periodismo-bases-datos-factiva-lexisnexis/>

Peter Morville & Jeffery Callender. *Search Patterns*. Beijing [etc.]: O'Reilly

Tony Russell-Rose, Tyler Tate (2013). *Designing the search experience. The Information Architecture of Discovery*. Amsterdam [etc.]: Elsevier

Weaver, D. A., & Bimber, B. (2008). Finding news stories: a comparison of searches using LexisNexis and Google News. *Journalism & Mass Communication Quarterly*, 85(3), 515-530

Wikipedia (2016). *Google News* [Online] https://en.wikipedia.org/wiki/Google_News

Abadal, Ernest (2002). "Elementos para la evaluación de interfaces de consulta de bases de datos". *El profesional de la información*, vol. 11, núm. 5 (septiembre-octubre), p. 349-360. <http://www.elprofesionalde lainformacion.com/contenidos/2002/septiembre/3.pdf>

Gracias