

# II Congreso Nacional I Congreso Internacional Hacia el Desarrollo Sostenible

**CTDS**  
Ciencia, Tecnología y Desarrollo Social



**SENNOVA**  
Sistema de Investigación,  
Desarrollo Tecnológico e Innovación

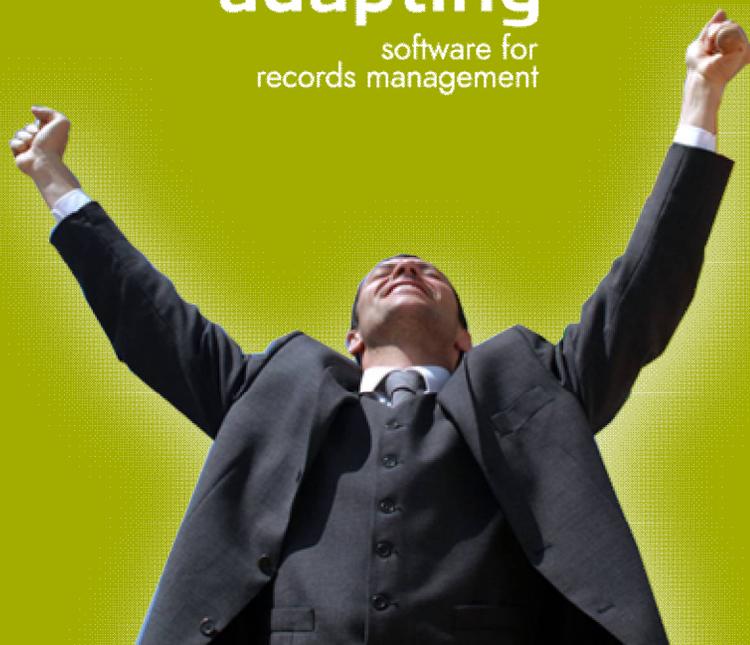
**grupo  
adapting**  
software for  
records management



## Search-Spider:

Proyecto de I+D con alcance internacional, apoyado por el Programa de Fomento de la Innovación y el Desarrollo Tecnológico en las empresas con enfoque Regional (2015-2017)

grupo  
**adapting**  
software for  
records management



1. **Quienes somos?**
2. **Participantes y plan de trabajo**
3. **Resultados en Colombia**
4. **Conclusiones y perspectivas**
5. **Demo de la aplicación**

grupo  
**adapting**  
software for  
records management



¿Quiénes somos?

¿Quién somos?



## Grupo Adapting:

Consultoría y desarrollo de soluciones de gestión documental profesional y ECM desde 1999.

>300 proyectos de gestión de contenidos, especializados en gestión documental.

Desde 2008, fabricante de SGD con red nacional e internacional de partners.

Sedes en: Valencia – Barranquilla - Bogotá.



## Adapting América S.A.S.

### **Sociedad colombiana:**

Creada en Barranquilla en marzo de 2011.

Oficina comercial en Bogotá.

Mesa de ayuda (call-center) en Barranquilla.

### 3 objetivos:

- 🗨 Soporte a partners en España/ Colombia
- 🗨 Desarrollo y localización de **abox**
- 🗨 Expansión a resto de Latinoamérica

**Visión:  
alcanzar  
liderazgo en  
SGD en  
Latinoamérica**



# Referencias principales en Colombia

- **Transmilenio S.A.:** proyecto Gestión Documental corp., BPM y CMS (CMIS 1.1), 2000 usuarios, empresa buses núm. 1. Partner: GTS.
- **Grupo Empresa Energía Bogotá:** gestor documental ciclo documental completo, EEB aprox. 250 usuarios (sector energía).
- **Eficacia:** gestor documental integrado con procesos clave de capital humano, facturación, aportes y compras (servicios). Partner: MTI.
- **Naves Colombia S.A.:** proceso administrativos del Grupo y conexión entre 8 sedes con solución abox (sector logístico).
- **Universidad Pedagógica y Tecnológica de Colombia (UPTC):** Ventanilla única y clasif. documental, aprox. 240 usuarios.
- **Lotería del Boyacá:** Ventanilla Única y procesos Dpto Jurídico, 40 usuarios activos (sector servicios).
- **Alcaldía de Monterrey (Casanare):** Correspondencia y clasif. exptes electrónicos. Gestión automat. PQRs, 40 usuarios (sector gobierno).
- **Fondecor:** fondo empleados de El Cerrejón para trámites asociados (sector asociativo).
- **Aguas Cartagena:** sistema de gestión documental con ECM para gestión de Expedientes de clientes, PQRs y contabilidad, 300 usuarios.
- **Concesión Ruta del Cacao:** Ventanilla única y procesos documentales.



grupo  
**adapting**  
software for  
records management



**Origen y objetivos**

## Planteamiento del problema



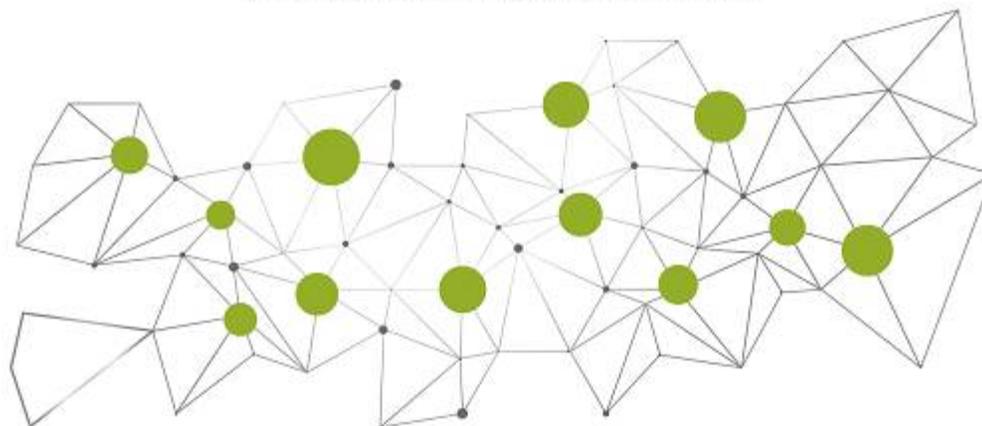
- 1) “Caos electrónico” aumenta exponencialmente.
- 2) El tiempo es ORO, sobre todo en entornos sanitarios.
- 3) Diversidad de repositorios: Estructurados vs. No estructurados + Internos vs. Externos
- 4) NO se aplican buenas prácticas de Gestión Documental y Archivística (p.ej. ISO-30300).
- 5) La seguridad es condición necesaria y emergencia social.
- 6) Buscadores “tontos”: NO consideran datos personales, preferencias, históricos + inteligencia colaborativa.

# Aplicación del Big Data e Inteligencia Artificial en búsqueda empresarial: piloto en sector Salud



Solución de búsqueda segura sobre repositorios de información heterogénea, basada en tecnologías de Inteligencia Artificial y Big Data, aprovechando la retroalimentación entre usuarios como valor diferencial.

Cofinanciado por: Fondo de Fomento de Innovación y Desarrollo Tecnológico del SENA.



# Objetivos

Desarrollar un **buscador** usable y seguro como SaaS, capaz de rastrear y extraer información de repositorios masivos de información heterogénea, mediante tecnologías de **Inteligencia Artificial** y **Big Data**, aprovechando la colaboración y la retroalimentación entre **usuarios** como valor diferencial.

## BÚSQUEDA EMPRESARIAL

- Acceso único a todas las fuentes de información del centro (captura integrada)
- Diseñar servidores terminológicos para hacer búsquedas semánticas
- Desarrollar una interfaz de buscador usable e integrada con red social privada

## BIG DATA E INTELIGENCIA ARTIFICIAL

- Emplear repositorios masivos y técnicas de rastreo e indexación de Big Data
- Aplicar técnicas de IA para optimizar las respuestas del sistema (priorizar contenidos)
- Aprovechar la inteligencia colaborativa en entornos corporativos

## SEGURIDAD

- Incorporar mecanismos de autenticación segura con doble factor.
- Integrar tecnologías de monitoreo en línea y detección de fraudes.

## Producto/ Servicio

- Instalable «on premises» / «cloud privado»
- Comercialización como producto independiente, tipo «plugin» que se enchufará a repositorios de la organización
- Requiere servicio de conexión/ parametrización inicial
- Se ofrecerá en pago-por-uso (p.ej. Núm. búsquedas procesadas).
- Se ofrecerá servicio de mejora continua de seguridad y de gestión de la información.



grupo  
**adapting**  
software for  
records management



**Participantes y  
plan de trabajo**

# Consorcio internacional Search-Spider

## COLOMBIA

- Adapting América (B/quilla) [www.adapting.com](http://www.adapting.com)
- ITI (subcontratado) [www.iti.es](http://www.iti.es)
- SENA (socio+beneficiario) [www.sena.edu.co](http://www.sena.edu.co)

**SIDICA**

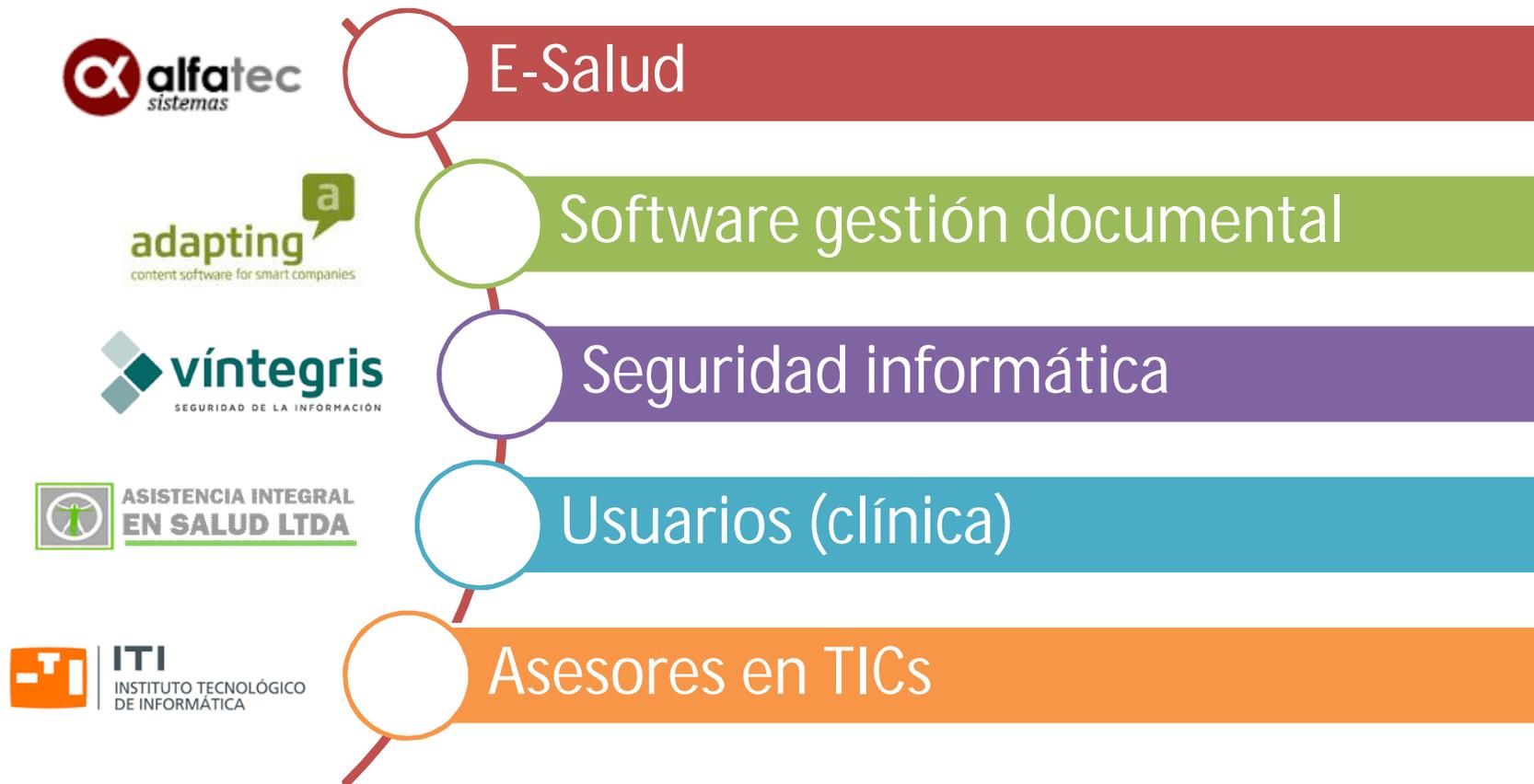


## ESPAÑA

- Alfatec Sistemas (Valencia) [www.alfatecsistemas.es](http://www.alfatecsistemas.es)
- Víntegris (Barcelona) [www.vintegris.com](http://www.vintegris.com)
- ITI (subcontratado) [www.iti.es](http://www.iti.es)



Desarrollo de tecnologías seguras para búsqueda en repositorios heterogéneos masivos de información sanitaria.



# Organismos públicos de apoyo a la I+D



Entidades de apoyo a la financiación de la I+D:

- ❑ Centro para el Desarrollo Tecnológico Industrial (**CDTI**, **España**)
- ❑ Convocatoria de Fomento de la Innovación y el Desarrollo Tecnológico en las Empresas, con enfoque Regional 2015-2017, Proceso II. Servicio Nacional de Aprendizaje (**SENA**, **Colombia**).

Presupuesto total (aprox.): 500.000 €





## Duración y fases del proyecto:

- Firma del Acuerdo de Consorcio (dic. 2016)
- Proyecto a **24 meses (España) / 12 meses (Colombia)**
- **4 fases:**
  - Análisis, requisitos y especificaciones
  - Desarrollos tecnológicos
  - Pilotos en España (1) y Colombia (2)
  - Cierre y conclusiones
- **Finales oficiales:**
  - España: Nov.-2018
  - Colombia: Dic.-2018



# Cronograma de actividades (Colombia)



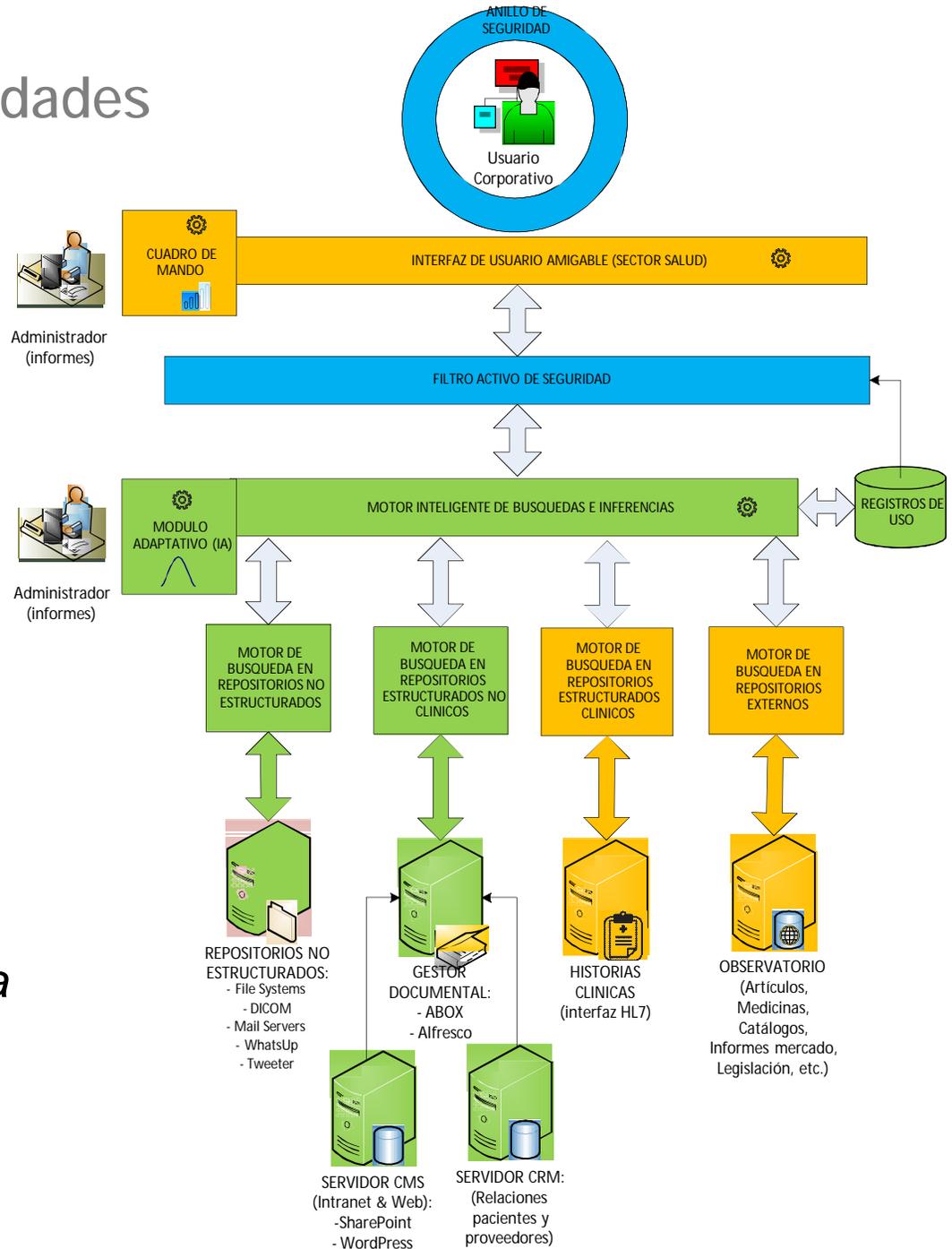
Número	Actividad	Desde	Hasta	Tiempo
<b>ACT01</b>	<b>ANÁLISIS DE REQUISITOS</b>	01/01/2018	28/02/2018	2 meses
ACT01-01	Análisis y especificación de requerimientos.	01/01/2018	28/02/2018	2 meses
ACT01-02	Diseño técnico y funcional. Acuerdo con piloto.	29/01/2018	28/02/2018	1 mes
<b>ACT02</b>	<b>DESARROLLOS TECNOLÓGICOS E INTEGRACIÓN</b>	01/03/2018	31/10/2018	8 meses
ACT02-01	<b>Desarrollo de los prototipos (5 componentes)</b>	01/03/2018	28/10/2018	8 meses
	Interfaz gráfica (frontend)	01/03/2018	30/04/2018	2 meses
	Servidor terminológico y ontologías	01/05/2018	31/05/2018	1 mes
	Motor de inferencias (inteligente y adaptativo)	01/06/2018	28/08/2018	3 meses
	Red social corporativa integrada (frontend)	29/08/2018	28/09/2018	1 mes
	Plataforma de indexación interoperable + Conectores fuentes diversas	29/09/2018	31/10/2018	1 mes
ACT02-02	Pruebas unitarias y pruebas de integración.	29/09/2018	31/10/2018	1 mes
<b>ACT03</b>	<b>PILOTO PARA VALIDACIÓN EN CENTRO DE SALUD DE COLOMBIA</b>	29/09/2018	28/01/2019	4 meses
ACT03-01	Instalación del piloto e integración de componentes externos (anillo seguridad y conector HCE)	29/09/2018	31/10/2018	1 mes
ACT03-02	Parametrización del software instalado en el piloto	29/09/2018	31/10/2018	1 mes
ACT03-03	Capacitación de usuarios y administradores	29/09/2018	31/10/2018	1 mes
ACT03-04	Operación de piloto y optimización de prototipos	29/10/2018	28/01/2019	3 meses
<b>ACT04</b>	<b>TRANSFERENCIA DE RESULTADOS AL SENA (CNCA+CCS)</b>	01/03/2018	28/01/2019	11 meses
	Inducción y participación de aprendices en proyecto	01/02/2018	30/09/2018	
ACT04-01	Curso-taller en tecnologías de Gestión Documental (Transferencia SENA)	01/04/2018	31/10/2018	7 meses
ACT04-02	Seminario de presentación y demostración del producto	01/12/2018	28/01/2019	2 meses
ACT04-03	Evento científico con participación de expertos nacionales e internacionales	30/12/2018	28/01/2019	1 mes

# Separación responsabilidades

SearchSpider



*Mínima dependencia cruzada*

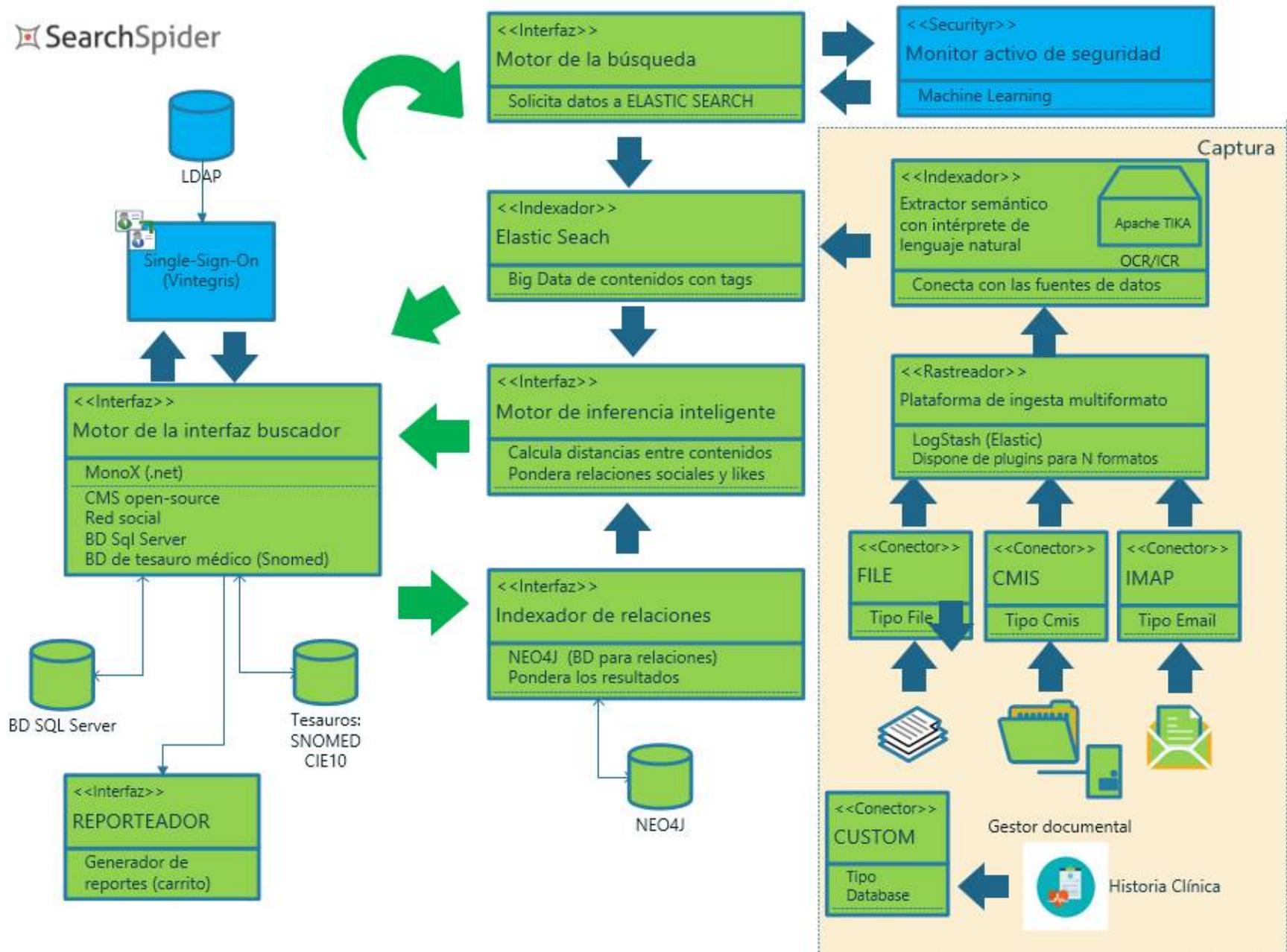


grupo  
**adapting**  
software for  
records management



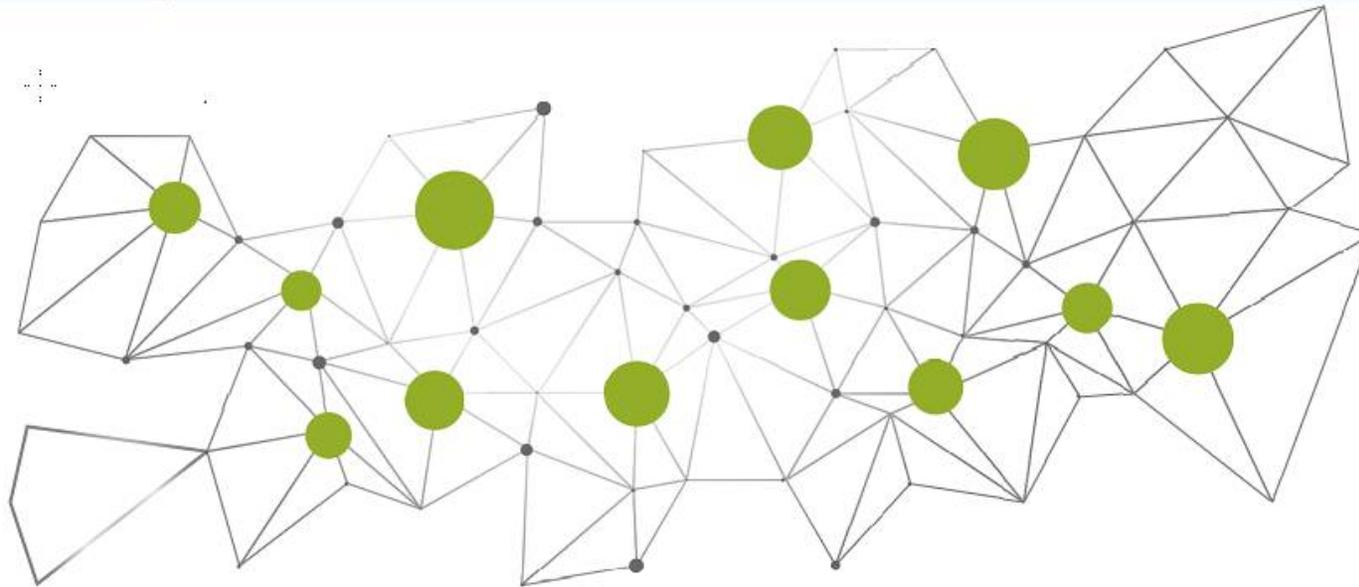
**Resultados en Colombia**

## DIAGRAMA DE ARQUITECTURA DEL BUSCADOR INTELIGENTE





# Módulos desarrollados



## Captura

- Rastreo automático en múltiples fuentes
- Captura en repositorios de texto, imágenes y correos electrónicos
- Captura en gestores documentales con enlace directo a contenidos
- Captura en fuentes externas (p.ej. RSS)
- Formato documental estándar (CMIS).

## Extracción

- Codificación por tesauros, p.ej. SNOMED
- Recuperación de sinónimos y abreviaturas
- OCR/ICR sobre cualquier imagen o documento
- Priorización de vocablos frecuentes o favoritos.

## Almacenamiento

- Big Data mediante plataforma Elastic
- Indexación de contenido de ficheros y de metadatos
- Métricas de distancia geométrica
- Almacenamiento y gestión 100% cloud.

## Buscador Inteligente

- Búsqueda semántica (por tesauros y relaciones)
- Búsqueda avanzada (por ocurrencia de texto)
- Inteligencia Artificial para priorizar resultados
- Asistente para generación de informes e histórico.

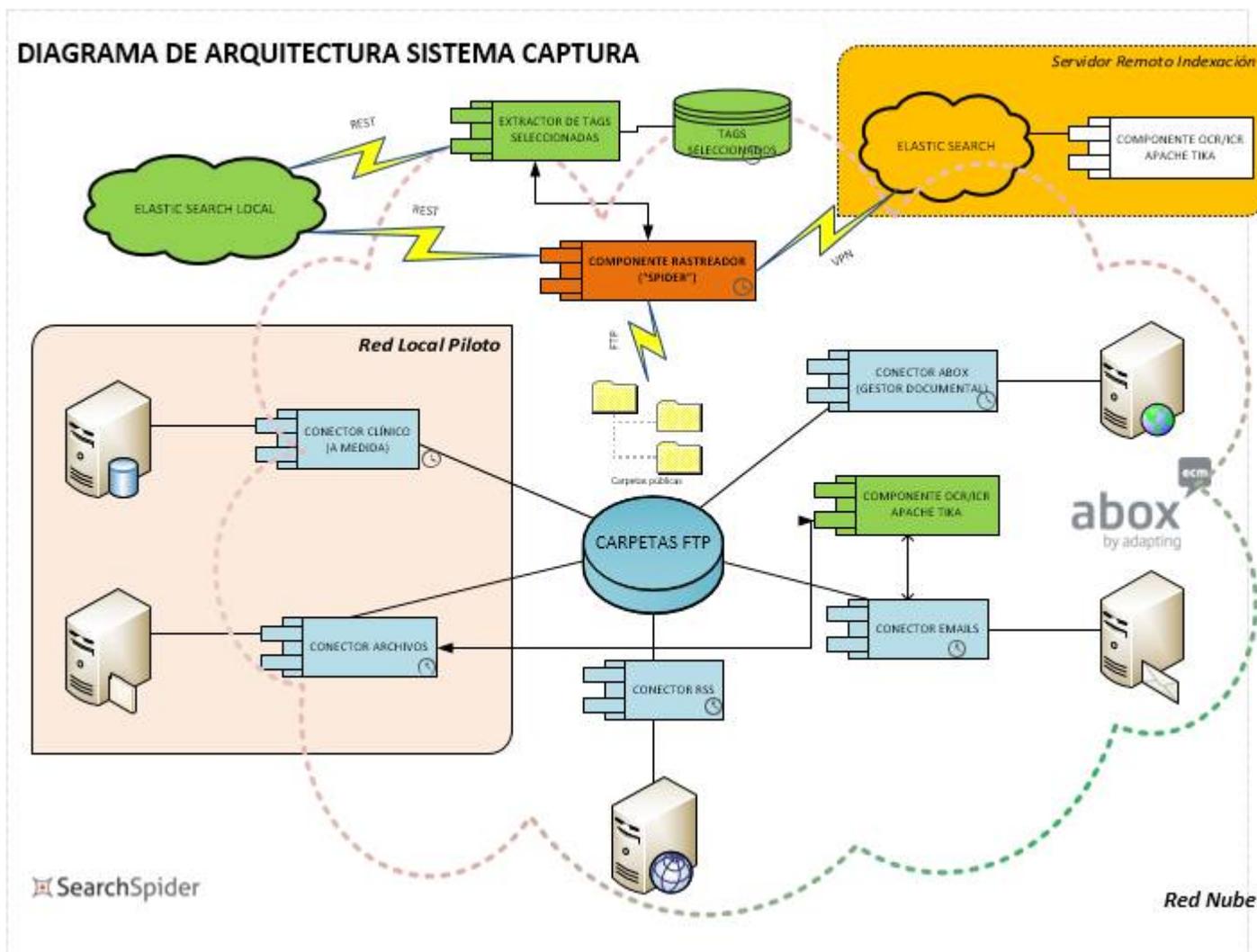
## Red Social Privada

- Posibilidad de comentar resultados y contenidos
- Declaración de afinidades entre usuarios
- Personalización del buscador
- Factor colaborativo para mejorar búsquedas (tecnología Neo4j).

## Seguridad Informática

- Login seguro con doble factor (confirmación por App móvil)
- Indexación de permisos en contenidos
- Motor inteligente para detección de comportamientos anómalos de usuarios .

# Módulo de captura



# Módulo de extracción

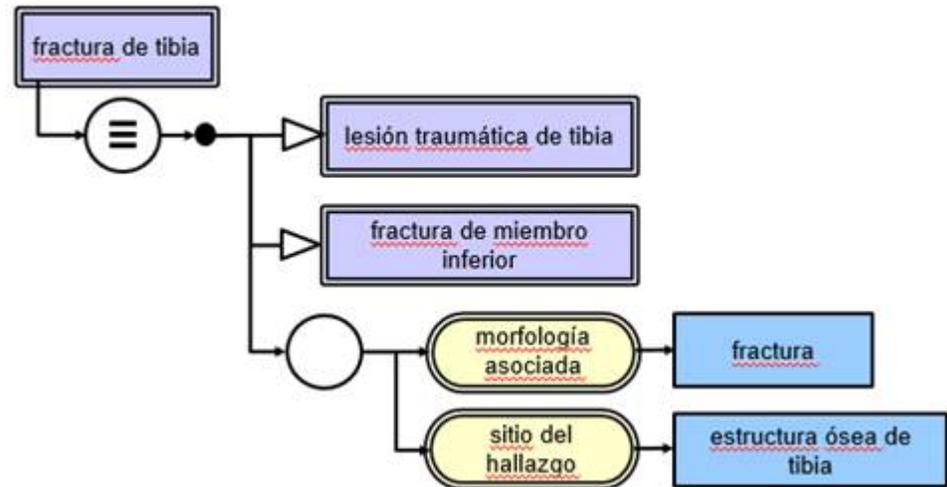


## Entradas:

- Tesauro médico aceptado (Snomed)
- Vocabulario priorizado (español)
- Sinónimos
- Abreviaturas
- OCR de imágenes

## Salidas:

- Tags de indexación por contenido
- Textos extraídos para consulta



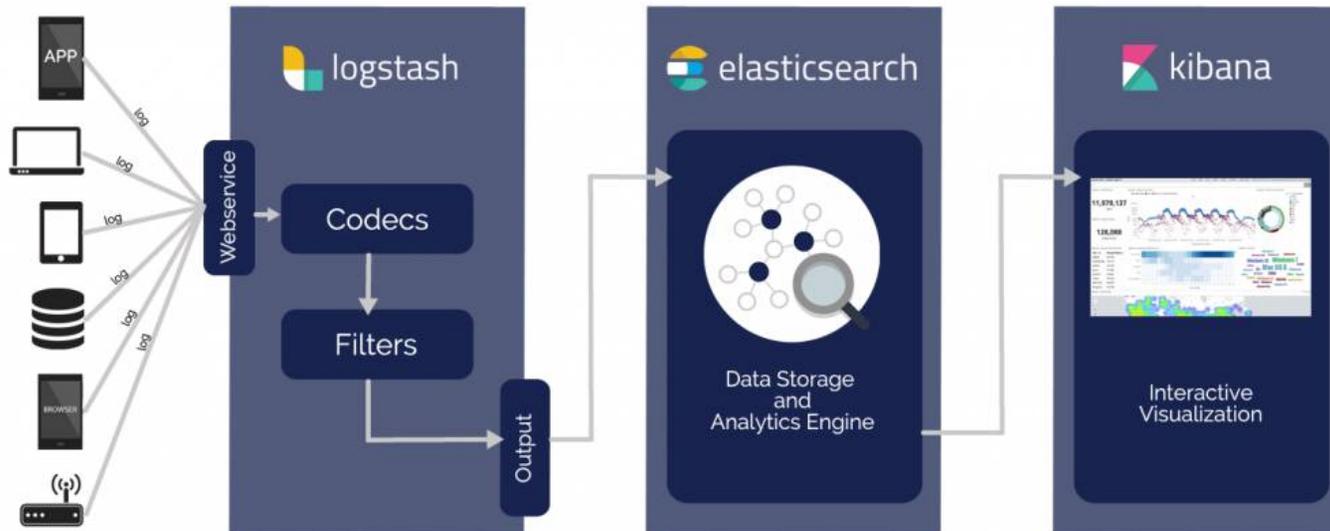
Leading healthcare terminology, worldwide

# Módulo de almacenamiento



## Nube de datos:

- Elastic Search
- Indexación de metadatos, permisos y contenido
- Instalación en la nube
- Métricas de similitud (“scoring”)
- Analítica de datos (Kibana)



# Buscador inteligente (i)



## Funciones:

- Búsqueda semántica (por tesauros y relaciones)
- Búsqueda avanzada (por ocurrencia de texto)
- Inteligencia Artificial para priorizar resultados
- Asistente para generación de informes e histórico

<https://search-spider.adapting.com>

The screenshot displays the SearchSpider web application interface. The top navigation bar includes links for 'Inicio', 'Mi muro', 'Preferencias', and 'Acerca de'. A search bar is prominently featured with the text 'dolor de m' and a dropdown menu showing related search suggestions such as 'presencia de dolor menstrual' and 'dolor al movimiento del útero'. The main content area shows search results for 'dolor leve', including a document titled 'ONCO\_2371293.TXT' with a star rating of 5. A sidebar on the right contains a 'Filtrar por:' section with input fields for 'Autor', 'Id. Paciente', and 'Nombre Archivo'. At the bottom right, a red-bordered box highlights a feedback section titled 'Opiniones sobre la búsqueda', which includes a star rating, a text input field for 'mi opinión sobre este resultado', and an 'Enviar Opinión' button. The interface is branded with 'ASISTENCIA INTEGRAL EN SALUD LTDA' and features a green and white color scheme.

# Buscador inteligente (ii)



<https://search-spider.adapting.com>

The screenshot displays the SearchSpider web application interface. At the top, there is a navigation bar with the SearchSpider logo and the text "ASISTENCIA INTEGRAL EN SALUD LTDA". The main header includes a search bar with the text "Buscar" and a magnifying glass icon. Below the search bar, there are filters for "medicina" and "general".

The search results are displayed in a list view. The first result is a PDF document titled "Introduccion a la medicina general integral.pdf" by "Null". The document details include:

- Datos CIP-Biblioteca Médica Nacional
- Sánchez Santos, Leonardo
- Introducción a la medicina general integral/ Leonardo Sánchez San
- Fecha de creación: 01/01/2018
- Fecha de Modificación: 01/01/2018
- Tags: medicina, integral, medicina general, general
- Descargar Archivo

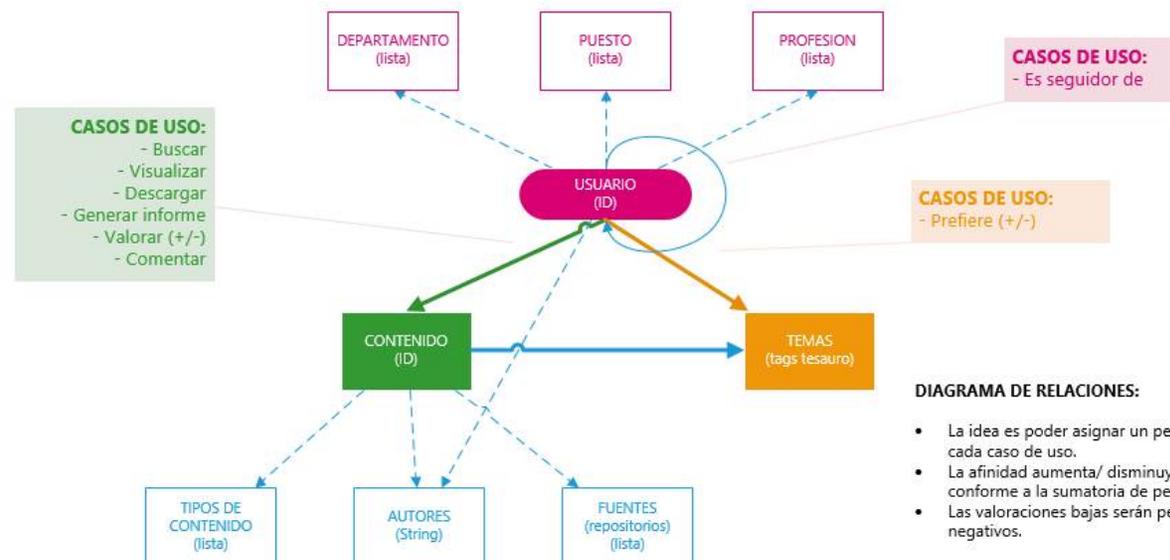
The interface also shows a sidebar with a user profile for "MonoX Admin" with 51 followers and a status "Su estado va aquí". The sidebar includes filters for "Filtrar por" and "Filtrar".

## Modelado de relaciones:

- Posibilidad de comentar resultados y contenidos
- Declaración de afinidades entre usuarios (seguidores y nivel de relación)
- Personalización del buscador
- Factor colaborativo para mejorar búsquedas (tecnología Neo4J)



### COMPONENTE DE INTELIGENCIA DE BÚSQUEDA



# Módulo de Inteligencia colaborativa (ii)

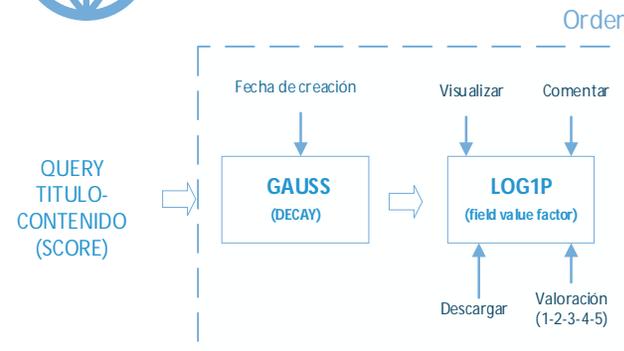


## Algoritmo de priorización resultados:

- Scoring de Elastic
- Rating del usuario (experiencia)
- Valoraciones usuarios (1-5)
- Novedad/ antigüedad del contenido
- Acciones realizadas por otros usuarios (pesos x acciones)
- Afinidades con otros usuarios (Neo4j)



ELASTIC SEARCH

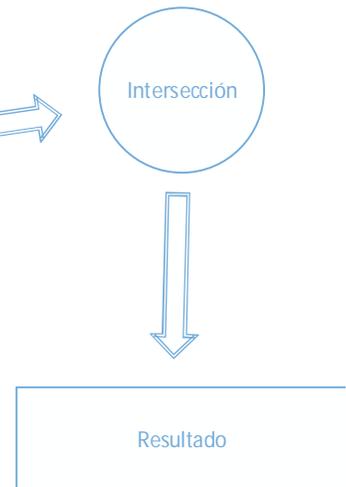
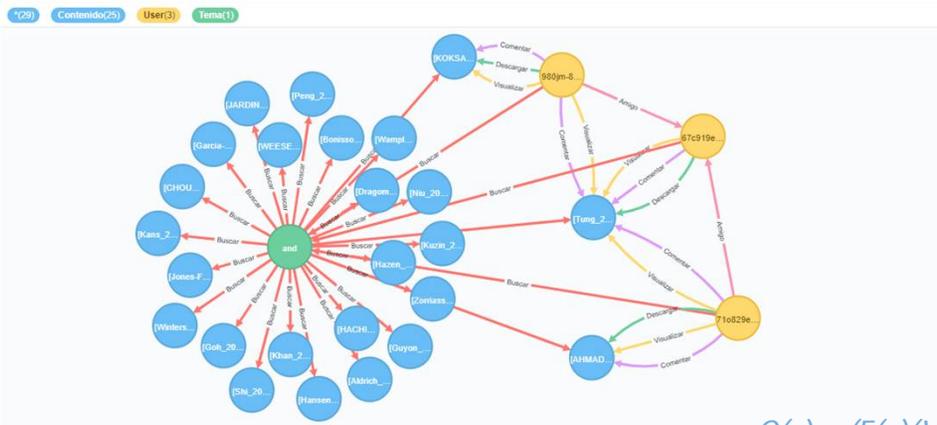


NEO4J



$$\sigma^2 = -scale^2 / (2 \cdot \ln(decay))$$

$$G(x) = \frac{(F(x)(Visualizar) + F(x)(Descargar) + F(x)(Comentar))}{3}$$



## Funciones:

- Login seguro con doble factor (confirmación por App móvil)
- Indexación de permisos en contenidos
- Motor inteligente para detección de comportamientos anómalos de usuarios



**NEBULASUITE®**  
Comprehensive **digital identity**  
protection for the Internet of Everything

**vintegrisTECH**

grupo  
**adapting**  
software for  
records management



**Conclusiones y perspectivas**

## Conclusiones:

 SearchSpider



### **Centro piloto del proyecto:**

- Licencia gratuita de Abox + Buscador Search-Spider (perpetuas)
- Tareas de adaptación e integración sufragadas por el proyecto.
- Mejora en la productividad interna

### **Para otros clientes potenciales (centros de salud, IPS, EPS):**

- Oportunidad de mejorar significativamente el comportamiento de los sistemas de gestión hospitalaria/ repositorios/ sistemas de información, sin grandes inversiones.
- Posibilidad de obtener información útil para la mejora de los procesos internos de gestión de la información y el conocimiento.

### **Para otros agentes:**

- Distribuidores de soluciones TIC (ampliación por servicio)
- Estudiantes, investigadores, profesionales (publicación nuevos conocimientos)
- SENA (experiencia de aprendices, transferencia a instructores, conocimientos para futuros aprendices).

# Perspectivas para Adapting:



## 1.) Integración de las nuevas tecnologías en producto:

- Nuevo buscador con Big Data integrado en **Abox-ECM**
- Nuevas oportunidades comerciales (“búsqueda corporativa”)

## 2) Reutilización de la tecnología para otras aplicaciones/ sectores:

- Sector jurídico-legal, sector educativo, etc.
- Aplicaciones basadas en “big data” sobre documentos

## 3) Nuevas iniciativas de I+D+i:

- Desarrollo de pronósticos basados en IA
- Reconocimiento y clasificación de patrones de comportamiento humano
- Open Data



Gracias por su atención



**Adapting América, S.A.S.**

Cra 48 No. 75-119 Oficina 205  
Barranquilla (Colombia)  
Tel. (5) 3605523  
america@adapting.com  
www.adapting.com

