risti
Revista Ibérica de Sistemas e Tecnologias de Informação
Revista Ibérica de Sistemas y Tecnologías de Información

Recebido/Submission: 24/03/2016
Aceitação/Acceptance: 27/04/2016

Recuperación de información a través del índice invertido en *Be Intelligent*

Sodel Vázquez-Reyes, María de León-Sigg, Perla Velasco-Elizondo,
Juan Villa- Cisneros, Sandra Briceño-Muro[1]

vazquezs@uaz.edu.mx, mleonsigg@uaz.edu.mx, pvelasco@uaz.edu.mx, jlvilla@uaz.edu.mx, sgbm0592@gmail.com

[1] Universidad Autónoma de Zacatecas, Ctra. Zacatecas-Guadalajara, Km. 6, Ejido La Escondida, 98160, Zacatecas, México.

DOI: 10.17013/risti.21.85–98

Resumen: Los avances en el almacenamiento de información han producido un aumento significativo de las colecciones de documentos digitales. La información contenida en estas colecciones es vital para la toma de decisiones, de manera que almacenar, indexar y recuperar información específica de estas colecciones es clave en las organizaciones, como lo establece el concepto de "búsqueda empresarial". Sin embargo, con la información almacenada en diferentes medios y formatos de archivo, la recuperación es una tarea compleja que los sistemas basados en palabras clave no resuelven completamente. Para proporcionar una solución a la administración de documentos digitales de *ADD Intelligence in Aviation*, se desarrolló el sistema *Be Intelligent* que utiliza un índice invertido para indexar y recuperar fácilmente los contenidos de documentos que satisfacen una consulta hecha en lenguaje natural. De esta manera se reduce el tiempo para encontrar la información que la empresa necesita durante la inspección de aeronaves.

Palabras-clave: índice invertido, recuperación de información, precisión, ranking recíproco promedio, búsqueda empresarial.

Information retrieval through inverted index in Be Intelligent system

Abstract: Advances in information storage have produced a significant rise of digital document collections. The information contained in these collections is vital to decision making, so store, index and recover specific information from these collections is key in organizations, as "enterprise search" concept establishes. However, with information stored in different media and file formats, retrieval is a complex task that systems based on keywords do not completely solve. To offer a solution to store, index and recover digital document contents for *ADD Intelligence in Aviation* enterprise, *Be Intelligent* system was developed. *Be Intelligent* uses an inverted index to easily index and recover the contents of documents that satisfy a natural language query. As a result, time to find information needed to inspect airplanes, is reduced.

Keywords: Inverted index, information retrieval, precision, mean reciprocal rank, enterprise search.

1. Introducción

Para mantenerse en competencia, las empresas necesitan tener acceso a grandes cantidades de información con la finalidad de mejorar y agilizar sus procesos de negocio. Sin embargo, los sistemas computacionales que recuperan documentos de una colección a través de palabras clave en una consulta de usuario, ya no son suficientes para satisfacer las necesidades de información, que generalmente están expresadas en lenguaje natural y que, por lo tanto, pueden ser semánticamente ambiguas.

En este contexto, ha aparecido el término de "búsqueda empresarial" (enterprise search), definido como la búsqueda en material digital de texto que pertenece a una organización, y que incluye la búsqueda en el sitio externo de la empresa, su intranet y en cualquier otro material electrónico de almacenamiento de texto como lo son los correos electrónicos, los registros en las bases de datos y los documentos compartidos (Stocker et al., 2014), (Hawking, 2004). La investigación en el campo de la búsqueda empresarial ha mostrado problemáticas que no pueden ser resueltas del todo con los sistemas de recuperación de información en base a palabras clave. Una de estas problemáticas tiene que ver con el tiempo para hacer búsquedas, que ocupa hasta el 25% del tiempo del trabajador que hace la búsqueda (Stocker, Richter, Kaiser, & Softic, 2015), y la segunda problemática trata con los aspectos relacionados con las implicaciones de la búsqueda manual de documentos que no siempre están disponibles para quien los necesita (Jadaan & Stenmark, 2008). Por ello, se espera que de alguna manera un sistema de recuperación de información sea capaz de interpretar los contenidos de un documento, extrayendo la información buscada pero también decidiendo la relevancia de la misma (Baeza-Yates & Ribeiro-Neto, 2011; Faria et al., 2015). Además, las estrategias utilizadas para aligerar procesos de software existentes en las pequeñas y medianas empresas deberían analizar los procesos para su adecuado aligeramiento, tal que, puedan obtenerse procesos optimizados que apoyen a la mejora continua de las organizaciones (Miramontes, Muñoz, Calvo-Manzano & Corona, 2016). En las pequeñas y medianas empresas la inversión en software si tiene relación directa con su direccionamiento estratégico, esto debido a que se percibe el valor agregado por el procesamiento de los datos que generan información relevante para la toma de decisiones (Riascos, Aguilera & Achicanoy 2016).

Esta necesidad ha transformado la recuperación de información en un importante campo de investigación y el desarrollo de sistemas computacionales que deben: *a)* ser capaces de procesar rápidamente grandes colecciones de documentos; *b)* permitir operaciones flexibles de búsqueda, y *c)* permitir la clasificación de la información recuperada (Manning, Raghavan, & Schütze, 2008). En este sentido, el propósito final de los sistemas de recuperación de información es ofrecer mecanismos que permitan a las empresas adquirir, producir y transmitir, al menor costo, datos e información con los atributos de calidad, precisión y validez, que sean útiles para la toma de decisiones (Arévalo, 2007). Sin embargo, es importante especificar que la función principal requerida de un sistema de recuperación de información no es devolver la información deseada por el usuario, sino indicar qué documentos son potencialmente relevantes para satisfacer su necesidad de información, porque, de hecho, un usuario de un sistema de recuperación de información está interesado en algún tema y no en los datos específicos que satisfacen una consulta. Los sistemas de recuperación de información tratan con texto, generalmente escrito en lenguaje natural, no bien estructurado y semánticamente ambiguo (Lara Navarra & Martínez Usero, 2006). En consecuencia, los documentos recuperados se consideran

útiles o no útiles, y la utilidad se juzga en términos de grados de efectividad, siendo la medida estándar la utilidad del documento recuperado (Blair, 2006).

En este documento se presenta el resultado de la investigación de búsqueda empresarial aplicado a *ADD Intelligence in Aviation*, una organización dedicada a la inspección de aeronaves, que dio como resultado el desarrollo de *Be Intelligent*, un sistema de recuperación de información que hace uso de índices invertidos para recuperar documentos relevantes para el personal de la empresa.

En la siguiente sección se describe la forma de trabajo de *ADD Intelligence in Aviation* previo al uso del sistema de recuperación de información. Posteriormente, se explica el uso de índices invertidos en sistemas de recuperación de información. En la sección cuatro se describen las métricas usadas para poder valorar la utilidad de los resultados del sistema *Be Intelligent*. En seguida, en la sección cinco se presentan los módulos principales del sistema *Be Intelligent*, así como las tecnologías utilizadas durante su construcción. La sección seis contiene la descripción de las pruebas realizadas a *Be Intelligent* para validar su utilidad en la empresa, y, finalmente, la sección siete contiene las conclusiones y los trabajos futuros del desarrollo que se presenta en este artículo.

2. ADD Intelligence in Aviation

ADD Intelligence in Aviation es un centro de ingeniería aeronáutica especializado en inspecciones físicas en aviones. En la actualidad, *ADD Intelligence in Aviation* trabaja con clientes distribuidos en ciudades mexicanas como Toluca, Ciudad de México y Sonora, pero también desarrolla proyectos en Seattle, WA, Santiago de Chile, en Chile, y Quetta, en Pakistán.

Cuando un proyecto de inspección comienza, los ingenieros de *ADD Intelligence in Aviation*, recopilan diferentes manuales, plantillas de inspección y otros documentos necesarios para verificar una aeronave asignada. Estos documentos se almacenan en diferentes formatos de archivos digitales, incluidos *.pdf, *.docx y *.xls. La colección de documentos se alimenta con manuales desarrollados por *ADD Intelligence in Aviation*, plantillas de inspección y directivas de aeronavegabilidad obtenidas en sitios web de diferentes agencias de aviación, como la Asociación de Transporte Aéreo Internacional, la Agencia Europea de Seguridad Aérea y el Mantenimiento de American Airlines. La colección de documentos se almacenaba en un disco duro externo perteneciente a la empresa. El acceso a este disco duro se realizaba a través de la red local de *ADD Intelligence in Aviation*, todo esto como parte de las políticas internas de la empresa, así como de los procesos de negocio de la organización. Esta situación implicaba que, si un ingeniero estaba trabajando fuera del alcance de la red local, realizaba una solicitud de alguna información que consideraba conveniente para el trabajo de inspección que estaba llevando a cabo. De manera local se hacía la búsqueda en el disco duro mencionado, se filtraban de manera manual los resultados de la misma, y se le enviaba al ingeniero solicitante una colección de documentos por correo electrónico o mediante un servicio de intercambio de archivos como Google Drive o Dropbox. Este proceso de recuperación de información implicaba tiempo de búsqueda y de envío, y dependía de la habilidad del personal que hacía el filtrado de manera manual de los documentos almacenados, de su conocimiento de los documentos existentes y de la organización del material almacenado. Por todo esto, se necesitaba un sistema que permitiera consultar

el contenido de los documentos almacenados en el disco duro, independientemente de la localización física del ingeniero que realizaba la consulta y que permitiera confirmar si ésta había sido útil para la inspección de aeronaves.

3. Estructura de datos de índice invertido para la recuperación de información.

Sabemos que la construcción de índices es necesaria para llevar a cabo búsquedas eficientes. Una estructura de datos llamada índice invertido, proporciona acceso a una lista de documentos que contienen el término buscado. El índice invertido es una lista de palabras y documentos en los que aparecen el término.

La mayoría de los sistemas de recuperación de información se basan en la estructura de datos de índice invertido. Esto permite acceder rápidamente a una lista de documentos que contienen un término junto con otra información (por ejemplo: el peso del término en cada documento, la posición relativa del término en cada documento, etc.). En recuperación de información, los objetos que son recuperados se denominan genéricamente "documentos". Dada una consulta proporcionada por el usuario, el motor de recuperación utiliza el índice invertido para ordenar por relevancia los documentos que contienen los términos de la consulta. Los términos que se consideran no informativos, son artículos, preposiciones y pronombres (el, en, de, a, etc.), llamados *stop-words*, y a menudo son ignorados.

El índice invertido explora el hecho de que, dada una consulta proporcionada por el usuario, la mayoría de los sistemas de recuperación de información sólo están interesados en mostrar un pequeño número de documentos que contienen algún término de la consulta. Debido a que todos los documentos están indexados por los términos que contienen, el proceso de generación, construcción y almacenamiento de las representaciones de documentos se denomina indexación y los archivos invertidos resultantes se denominan índice invertido. La construcción de un índice invertido para mantener cualquier tipo de sistema de búsqueda requiere realizar una serie de pasos: almacenar los documentos, eliminar las *stop-words* y, finalmente, fusionar y almacenar los términos en el índice invertido (Manning et al., 2008). El proceso de construcción del índice o indexación se muestra en la Figura 1.

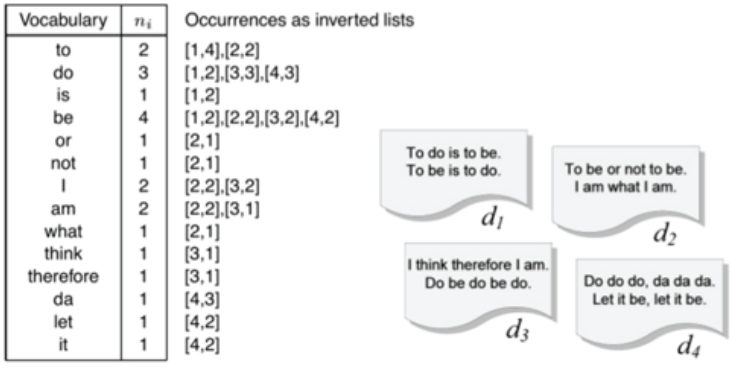

Figura 1 – Proceso de construcción del índice (Martín-Daucasa, 2012) .

4. Métricas para los documentos recuperados por un sistema de recuperación de información

Como la utilidad es un criterio complejo para medir, porque se basa principalmente en el juicio de alguien (usuario o no usuario) (Arquero Avilés & Salvador Oliván, 2004), se utilizan comúnmente varias métricas. Ejemplos de éstas son la *Precisión*, el *Ranking Recíproco* y el *Ranking Recíproco Promedio*. La *Precisión* (1) se define como la proporción de documentos relevantes recuperados (Martínez Méndez, 2004). Esta métrica evalúa la capacidad del sistema para posicionar primero la mayoría de los documentos relevantes y mide el porcentaje de documentos recuperados que tienen relevancia (Martínez Méndez, 2004). Su cálculo se obtiene con

$$Precisión = \frac{documentos\ relevantes\ recuperados}{documentos\ recuperados} \quad (1)$$

Por otro lado, el *Ranking Recíproco - R*R (2), se utiliza para medir la capacidad del sistema para recuperar documentos relevantes en las posiciones más altas en la lista de resultados. Se calcula mediante la siguiente ecuación (Levene, 2010):

$$RR = \frac{1}{ranking(i)} \quad (2)$$

Donde *ranking(i)* se refiere a la posición del documento que contiene la información correcta para la consulta *i*, y *RR* será *cero* si la información buscada no se encuentra en ningún documento.

Por último, el *Ranking Recíproco Promedio - MRR* (3) es el promedio de los valores *RR* para todas las consultas (Levene, 2010). Esta métrica da la puntuación más alta a los documentos que se encuentran en las primeras posiciones de la lista de resultados, ya que mide la precisión y el orden de los resultados correctos (Levene, 2010). Esta métrica se calcula con:

$$MMR = \frac{\sum_{i=1}^{|Q|} RR}{|Q|} \quad (3)$$

Donde *RR* es el *Rankling Recíproco*, mostrado en (2), y *Q* es el número de consultas. Para facilitar la recuperación de documentos o partes de documentos, se crean índices. De esta forma, los índices, proporcionan a los usuarios medios efectivos y sistemáticos para ubicar documentos relevantes que satisfagan las necesidades o peticiones de información (Anderson, 1997). Hay varias estructuras de índice, un índice invertido es un mecanismo orientado a palabras formadas por dos elementos: 1) el vocabulario, definido como el conjunto de diferentes términos (palabras) en los textos; y 2) las listas de ocurrencias, definidas como la lista de documentos en los que aparece un término determinado (Manning et al., 2008).

5. Sistema *Be Intelligent*.

El sistema *Be Intelligent* contiene dos módulos principales: uno para el almacenamiento e indexación de la información, como se puede observar en la Figura 2, y uno para la recuperación de documentos, que se muestra en la Figura 3. El primer módulo (Figura 2), permite almacenar los documentos en un servidor y enviar datos, metadatos y contenido de los documentos al índice. A este módulo solamente tienen acceso los administradores del sistema, que son quienes previamente reúnen y valoran los documentos que serán almacenados e indexados. El sistema soporta formatos de archivos digitales .pdf, .xls, .xlsx, .txt, .doc y .docx. Los datos a capturar, y que son enviados al servicio de indexación, son: título del documento, autor del documento, número de páginas, año de publicación, imagen descriptiva para la identificación del documento y archivo que contiene el documento. Esta información es la que se utiliza para generar el índice que permite la recuperación del documento.

Para la construcción de este módulo se utilizó Apache Solr, que provee el soporte para la indexación. Esta herramienta de código libre es una API escrita en Java que es capaz de realizar consultas mediante frases y búsquedas por proximidad, y que actualiza el índice simultáneamente.

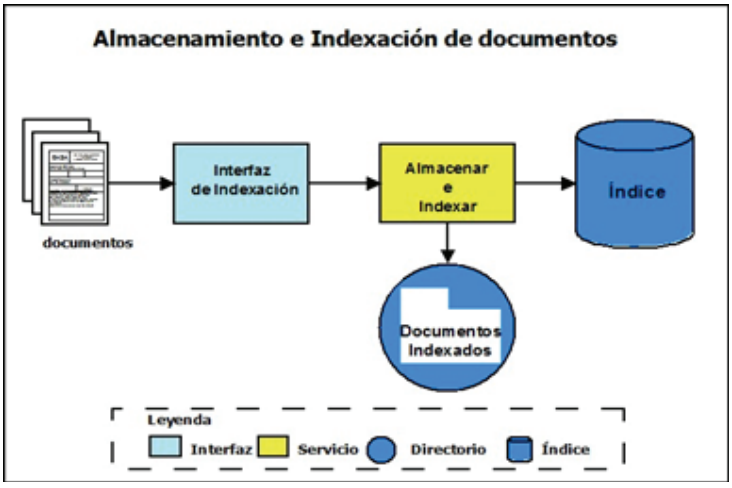

Figura 2 – Módulo de almacenamiento e indexación de documentos

El módulo de recuperación de información recibe una consulta en lenguaje natural y solicita al motor de recuperación de información los documentos que coinciden con la consulta (Figura 3). Este módulo tiene dos elementos fundamentales, el primero es la interfaz de búsqueda, mediante la cual los usuarios ingresan las consultas al sistema en lenguaje natural, y la segunda es el servicio de búsqueda, que procesa las solicitudes de información y muestra los resultados obtenidos durante la consulta en orden de relevancia.

El componente de interfaz contiene un formulario simple desarrollado en HTML5. El servicio de búsqueda se realiza a través de peticiones por http GET

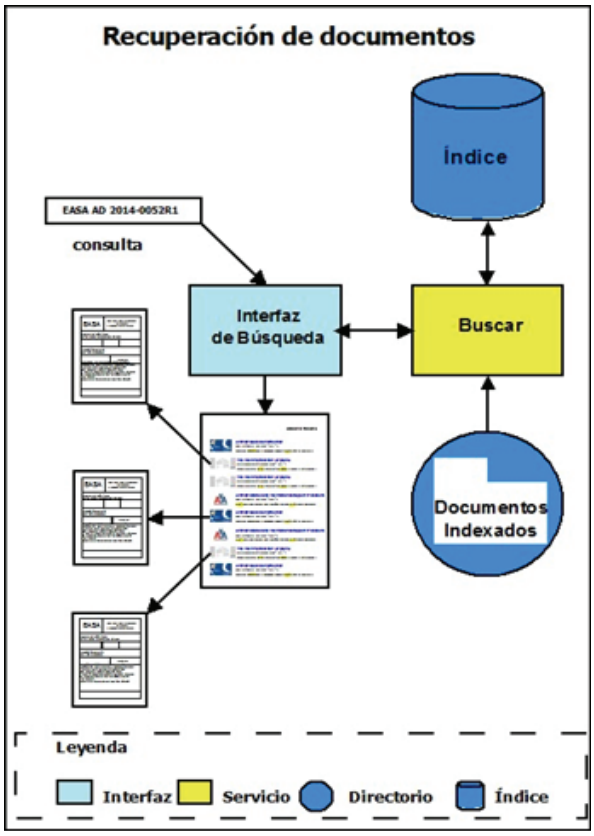

Figura 3 – Módulo de recuperación de documentos

6. Evaluación al sistema *Be Intelligent*.

Como se describió en la Sección 4, existen varias métricas para evaluar un sistema de recuperación de información. En esta sección se presentan los valores obtenidos para esas métricas con el sistema *Be Intelligent*.

Para evaluar el sistema *Be Intelligent*, creó una colección. La colección contiene 300 documentos: 100 de ellos con información relacionada con el control, mantenimiento e inspecciones de aeronaves; 115 documentos con información sobre ingeniería de software y 85 documentos con información para el desarrollo de investigación y la recuperación de información. Además, se crearon dos grupos de usuarios, cada uno con cinco usuarios de prueba, como se describe a continuación:

Grupo 1: Estudiantes de ingeniería de software, sin experiencia en sistemas de recuperación de información, sin una necesidad específica de información.

Grupo 2: Ingenieros de *ADD Intelligence in Aviation*, con necesidades específicas de información para realizar las inspecciones de aeronaves asignadas.

Las pruebas se organizaron en dos fases. La primera fase corresponde a una consulta breve; la fase dos corresponde a consultas largas. Esta dos fases permite la evaluación en la recuperación de documentos relevantes obtenidos por el sistema *Be Intelligent*. Cada fase consistió en los siguientes pasos: a) acceso a la opción *Discover* en el sistema *Be Intelligent*; b) búsqueda de un concepto o una frase dependiendo de la fase (en la Figura 4 se muestra una captura de pantalla del sistema para la búsqueda con el término *airplane*); c) revisión de los documentos recuperados; d) registro de la cuenta del total de documentos encontrados, el número total de documentos relevantes identificados y la posición del documento recuperado más relevante, todo esto en un formulario web después de cada consulta.

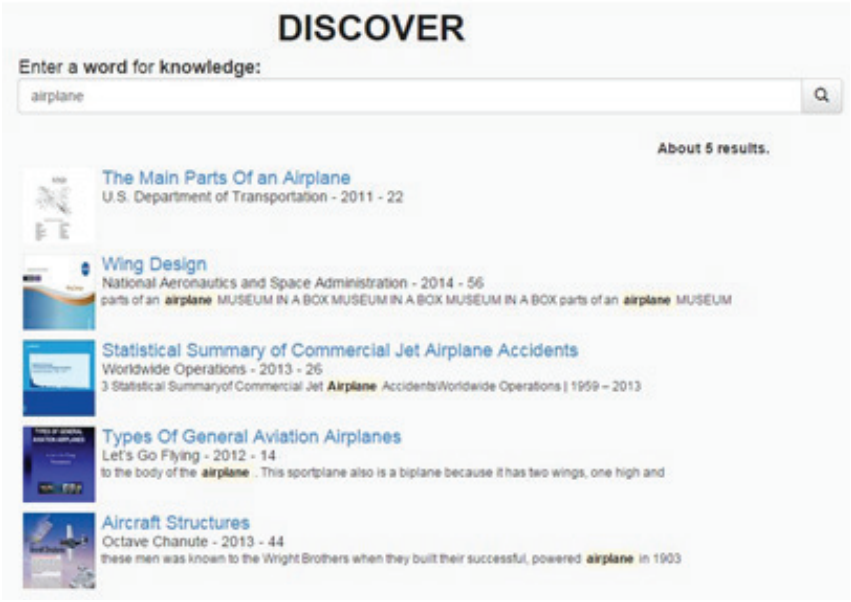

Figura 4 – Captura de pantalla de los documentos recuperados por el sistema *Be Intelligent* cuando se busca el término "airplane".

Los resultados obtenidos para cada grupo y cada fase (descritos en la Sección 6), se muestran a continuación.

6.1. Resultados del grupo 1.

Fase 1. La Figura 5 muestra la *Precisión* calculada para el Grupo 1-Fase 1. Ninguna búsqueda obtuvo 1.0 y la *Precisión* promedio fue de 0.73. Los mejores resultados fueron obtenidos por los usuarios de las prueba tres y cinco, con un valor de 0.78. Este valor se considera aceptable porque el dominio de la consulta para este grupo fue más amplio.

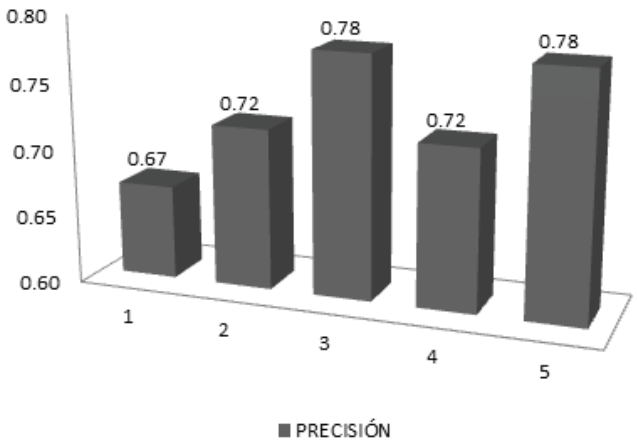

Figura 5 – *Precisión* calculada para el Grupo 1 - Fase 1.

El *Ranking Recíproco Promedio (MRR)* calculado para las consultas del Grupo 1-Fase 1 se muestra en la Tabla 1. En esta tabla se observa que el primer documento relevante se ubicó en la tercera posición, en las consultas tres y cuatro. Con estos datos, *MRR* fue de 0.27. Este valor estuvo lejos del valor ideal de 1.0, pero debería considerarse que el nivel de experiencia del grupo es bajo por lo que sus consultas eran menos específicas y el dominio de la consulta era más amplio. Debido a esto, *Be Intelligent* respondió adecuadamente en este contexto.

Usuario	Ranking	RR
1	4	1/4 = 0.25
2	5	1/5 = 0.20
3	3	1/3 = 0.33
4	3	1/3 = 0.33
5	4	1/4 = 0.25
	MRR	= 0.27

Tabla 1 – *Ranking Recíproco Promedio* calculado con los resultados del Grupo 1-Fase 1.

Fase 2. Los resultados obtenidos con una consulta larga realizada por el Grupo 1-Fase 2, se muestran en la Figura 6. En esta figura se puede observar que los usuarios uno y dos obtuvieron una *Precisión* de 1.0, mientras que el valor promedio de *Precisión* es 0.89, que está muy cerca del valor ideal de 1.0.

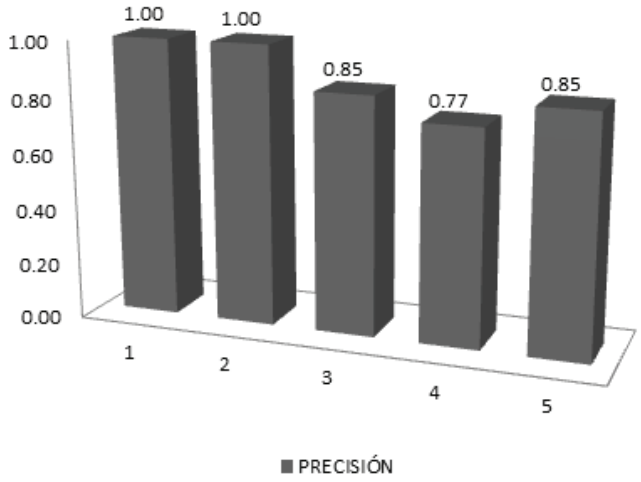

Figura 6 – *Precisión* calculada para el Grupo 1- Fase 2

El *MRR* para el documento recuperado en esta fase fue 0.29, como se muestra en la Tabla 2. Los resultados con los documentos relevantes mejor posicionados fueron las consultas uno, tres y cinco, donde el primer documento relevante se encontró en la tercera posición. El valor *MRR* es el esperado porque el sistema no responde a consultas largas con un grupo de un bajo nivel de experiencia.

Usuario	Ranking	RR
1	3	1/3 = 0.33
2	5	1/5 = 0.20
3	3	1/3 = 0.33
4	4	1/4 = 0.25
5	3	1/3 = 0.33
	MRR	= 0.29

Tabla 2 – *Ranking Recíproco Promedio* calculado con los resultados del Grupo 1-Fase 2

6.2. Resultados del Grupo 2

En esta sección se muestran los resultados obtenidos con el Grupo 2. Los ingenieros de *ADD Intelligence in Aviation* formaron este grupo, y por ello su nivel de experiencia con los sistemas de búsqueda es alto, así como sus necesidades de información son claras.

Fase 1. En esta fase, la consulta establecida fue la directiva *FAA-2013-0695*, un término muy específico. La *precisión* promedio fue de 0.67, y los usuarios tres y cuatro obtuvieron un valor de 1.0 de *precisión*. Sin embargo, los usuarios dos y cinco obtuvieron

valores de *precisión* por debajo del promedio porque necesitaban más detalles sobre la directiva, como se muestra en la Figura 7. De acuerdo con estos resultados, un 67% de los documentos recuperados fueron relevantes, debido a esto, el sistema respondió adecuadamente durante esta parte del experimento.

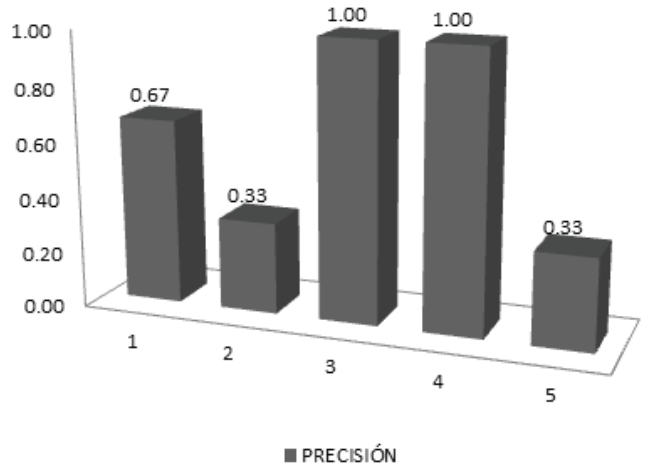

Figura 7 – La precisión calculada para el Grupo 2-Fase 1.

A partir de los datos presentados en la Tabla 3, se muestra que *MRR* es 1.0, porque todos los usuarios de la prueba indicaron que el primer documento mostrado era el más relevante para esta consulta. El primer documento recuperado fue juzgado como útil; por esta razón, el sistema se juzga como efectivo.

Usuario	Ranking	RR
1	1	1/1 = 1
2	1	1/1 = 1
3	1	1/1 = 1
4	1	1/1 = 1
5	1	1/1 = 1
	MRR	= 1.0

Tabla 3 – *Ranking Recíproco Promedio* calculado con los resultados del Grupo 2-Fase 1

Fase 2. Los datos se muestran en la Figura 8. La *precisión* promedio para esta fase del experimento fue de 0.83, lo que indica que más del 80% de los documentos recuperados son relevantes, mejorando la toma de decisiones del ingeniero.

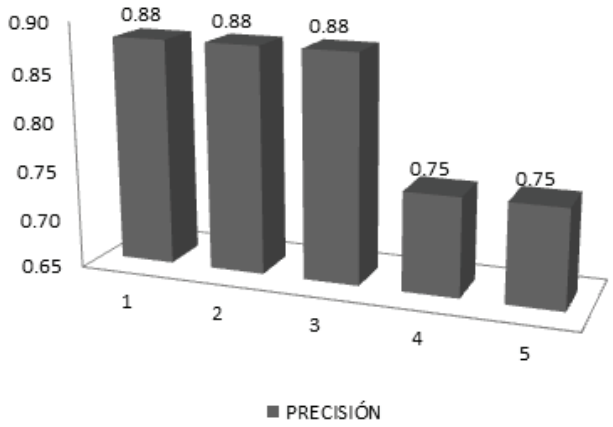

Figura 8 – *Precisión* calculada para el Grupo 2 Fase 2

En la Tabla 4, se muestran los cálculos *RR* y *MRR*. Como puede observarse, en el 80% de las consultas el documento con la información más relevante para la consulta se encuentra en la primera posición. Debido a esto, el ingeniero encontró información relevante sin perder demasiado tiempo revisando otros documentos no relevantes.

Usuario	Ranking	*RR*
1	1	1⁄1 = 1
2	1	1⁄1 = 1
3	2	1⁄2 = 0.5
4	1	1⁄1 = 1
5	2	1⁄2 = 0.5
	MRR	= 0.80

Tabla 4 – *Ranking Recíproco Promedio* calculado con los resultados del Grupo 2-Fase 2

7. Conclusiones

Los sistemas computacionales utilizados para recuperar información a partir de una gran colección de documentos deben ser capaces de reducir el tiempo de búsqueda y permitir la clasificación de la información recuperada en orden de relevancia al menor costo, pero con alta calidad y precisión para ser útiles en la toma de decisiones. *ADD Intelligence in Aviation* necesitaba tal sistema, y en esta investigación, se mostraron los resultados de la implementación de una estructura de datos de índice invertido en el sistema *Be Intelligent*. Se pueden identificar dos contribuciones principales del sistema *Be Intelligent*. En primer lugar, la creación de un esquema de almacenamiento e indexación que facilita la gestión de la información de documentos de uso rutinario

en una empresa. Además el esquema de indexación permite incluir de forma cómoda nuevos documentos en la estructura de indexación, la esencia del sistema *Be Intelligent*.

En segundo lugar, se proporcionó un método fácil, rápido y preciso para acceder a su colección de documentos a través de consultas expresadas en lenguaje natural. Los usuarios ahora pueden recuperar la información necesaria para realizar la inspección de aeronaves, sin la necesidad de hacer una búsqueda en un disco duro externo y sin invertir demasiado tiempo revisando documentos no relevantes.

De esta manera, la aplicación de índice invertido está ayudando a la empresa para cumplir sus deberes de una forma más productiva.

Podemos describir como un trabajo futuro mejorar la recuperación de información usando un algoritmo con un enfoque de recuperación de pasajes. Podría consistir en cuatro puntos: 1.- Ejecutar una recuperación completa de documentos, 2. Dividir los documentos recuperados en pasajes, 3.- Ejecutar la recuperación de pasaje contra el conjunto de pasajes creado en los puntos 2, y 4.- Listar los pasajes relevantes recuperados ordenados por relevancia. Para decidir si una estrategia de recuperación de pasajes es útil o no, es necesario evaluar su capacidad para recuperar pasajes de forma eficiente.

Referencias

Anderson, J. D. (1997). *Guidelines for Indexes and Related Information Retrieval Devices*. Bethesda, MD.

Arévalo, J. A. (2007). Gestión de la Información, Gestión de Contenidos y Conocimiento. In *II Jornadas de Trabajo del Grupo SIOU* (pp. 1–15). Universidad de Salmanaca.

Arquero Avilés, R., & Salvador Oliván, J. A. (2004). La Investigación en Recuperación de Información: Revisión de Tendencias Actuales y Críticas. *Cuadernos de Documentación Multimedia*, (15), 2–3.

Baeza-Yates, R., & Ribeiro-Neto, B. (2011). *Modern Information Retrieval: The Concepts and Technology behind Search* (2nd. editi). Addison-Wesley Professional.

Blair, D. C. (2006). The data-document distinction revisited. *ACM SIGMIS Database*, *37*(1), 77–96. DOI: 10.1145/1120501.1120507

Faria, B. M., Gonçalves, J., Reis, L. P., & Rocha, Á. (2015). A Clinical Support System Based on Quality of Life Estimation. *Journal of Medical Systems*, 39(10), 114. DOI: 10.1007/s10916-015-0308-1

Hawking, D. (2004). Challenges in Enterprise Search. In *Proceedings of the 15th Australasian database conference* (pp. 15–25). Australian Computer Society, Inc.

Jadaan, T., & Stenmark, D. (2008). Knowledge Worker's Use of Electronic Resources. In *Proceedings of the 16th European Conference on Information Systems*. Galway, Ireland.

Lara Navarra, P., & Martínez Usero, J. A. (2006). *Agentes Inteligentes en la Búsqueda y Recuperación de Información* (Segunda Ed). Barcelona, España: Planeta-UOC, S. L.

Levene, M. (2010). *An Introduction to Search Engines and Web Navigation*. Wiley Publishing, Inc.

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press.

Martín-Daucasa, C. (2012). *Desing and Evaluation of new XML Retrieval Methods and their Application to Parliamentary Documents*. Universidad de Granada.

Martínez Méndez, F. J. (2004). *Recuperación de información: Modelos, Sistemas y Evaluación*. Murcia, España: KIOSKO JMC.

Miramontes, J., Muñoz, M., Calvo-Manzano, J., & Corona, B. (2016) Establecimiento del estado del arte sobre el aligeramiento de procesos de software. *Revista Ibérica de Sistemas Y Tecnologías de La Información,* DOI: 10.17013/risti.17.16–25

Riascos, S., Aguilera, A., & Achicanoy, H. (2016) Inversión en Tecnologías de la Información y las Comunicaciones y su relación con en el direccionamiento estratégico de las PYMES de Santiago de Cali - Colombia. *Revista Ibérica de Sistemas Y Tecnologías de La Información,* DOI: 10.17013/risti.18.1–17

Stocker, A., Richter, A., Kaiser, C., & Softic, S. (2015). Exploring Barriers of Enterprise Search Implementation: a Qualitative User Study. *Aslib Journal of Information Management*, *67*(5), 470–491.

Stocker, A., Zoier, M., Softic, S., Paschke, S., Bischofter, H., & Kern, R. (2014). Is Enterprise Search Useful At All? Lessons Learned From Studying User Behavior. *Proceedings of the 14th International Conference on Knowledge Technologies and Data-Driven Business*, (January). DOI: 10.1145/2637748.2638425