

**SISTEMA MULTIAGENTE PARA LA OPTIMIZACION DE  
BUSQUEDAS EN EL DOMINIO MAMPOSTERIA ESTRUCTURAL  
BASADO EN TECNICAS DE WEB SEMANTICA**



**OLGA LUCIA RODRIGUEZ MENESES  
CAROL ROSANNA LOAIZA SALAZAR**

**Trabajo de grado**

**Director: Ing. Juan Carlos Vidal Rojas**

**UNIVERSIDAD DEL CAUCA  
FACULTAD DE INGENIERIA ELECTRONICA Y  
TELECOMUNICACIONES, DEPARTAMENTO DE SISTEMAS  
POPAYAN, ABRIL DE 2005**

<b>1</b>	<b>ESTADO DEL ARTE: EVOLUCIÓN DE LA WEB ACTUAL HACIA LA WEB SEMÁNTICA.</b>	<b>9</b>
1.1	WEB SEMÁNTICA.....	9
1.2	ONTOLOGIAS EN WEB SEMANTICA.....	10
1.3	SISTEMAS MULTIAGENTE EN LA RECUPERACION DE INFORMACION EN LA WEB.....	10
1.4	MODELO MATEMATICO PARA RECUPERACION DE INFORMACION.....	13
1.4.1	Modelo Vectorial.....	13
1.5	TRABAJOS RELACIONADOS.....	16
<b>2</b>	<b>MODELO DEL DOMINIO</b>	<b>19</b>
2.1	CRITERIOS DE DISEÑO DE ONTOLOGIAS.....	20
2.1.1	Principios de Diseño.....	20
2.2	METODOLOGÍA ADAPTADA.....	22
2.2.1	Especificación.....	22
2.2.2	Adquisición del conocimiento.....	22
2.2.3	Conceptualización.....	23
2.2.4	Implementación.....	26
2.2.5	Evaluación.....	27
2.2.6	Documentación.....	27
2.3	APLICACIÓN DE LA METODOLOGÍA ADAPTADA EN LA CONSTRUCCIÓN DE LA ONTOLOGÍA.....	27
2.3.1	Documento de Especificación de Requerimientos de la Ontología DERO.....	27
2.3.2	Conceptualización.....	29
2.3.3	Implementación.....	34
2.3.4	Documentación.....	34
<b>3.</b>	<b>SISTEMA MULTIAGENTE (MULTI-AGENT SYSTEM MAS)</b>	<b>35</b>
3.1	PROPÓSITO Y COMPONENTES DEL SISTEMA.....	36
3.1.1	Arquitectura del sistema.....	36
3.2	PROCESO DE DESARROLLO DEL SISTEMA MULTIAGENTE (MAS).....	37
3.2.1	Especificación de requerimientos.....	38
3.2.1.1	Modelo de Casos de Uso.....	39
3.2.2	Análisis.....	42
3.2.3	Diseño.....	57
3.2.4	Implementación.....	66
<b>4.</b>	<b>HERRAMIENTAS TECNOLÓGICAS EMPLEADAS PARA LA IMPLEMENTACIÓN DEL SISTEMA</b>	<b>73</b>
4.1	TECNOLOGIA JAVA.....	73

<b>4.2 PLATAFORMA PARA LA IMPLEMENTACION DE SISTEMAS MULTIAGENTE.....</b>	<b>74</b>
4.2.1 JADE (Java Agent DEvelopment Framework).....	74
<b>4.3 TECNOLOGIA PARA LA MANIPULACION DE ONTOLOGIAS. ....</b>	<b>74</b>
4.3.1 Protégé OWL Plugin.....	74
<b>4.4 HERRAMIENTA PARA LA EJECUCION DE CONSULTAS EN INTERNET .....</b>	<b>75</b>
4.4.1 Google APIs Service.....	75
<b>4.5 SERVIDOR DE PÁGINAS WEB.....</b>	<b>75</b>
4.5.1 Tomcat.....	75
<b>5 PROTOTIPO DE PORTAL WEB</b>	<b>76</b>
<b>6 PRUEBAS REALIZADAS</b>	<b>77</b>
<b>7 CONCLUSIONES</b>	<b>114</b>
<b>8. TRABAJO FUTURO</b>	<b>115</b>
<b>9 REFERENCIAS</b>	<b>116</b>

## INTRODUCCION

Internet ha supuesto una revolución sin precedentes en el mundo de la informática y de las comunicaciones; los inventos del telégrafo, teléfono, radio y computador sentaron las bases para esta integración de capacidades nunca antes vivida [TOR, 1998]. Internet es a la vez una oportunidad de difusión mundial, un mecanismo de propagación de la información y un medio de colaboración e interacción entre individuos y computadores independientemente de su localización geográfica [BAR, 1999].

Hasta hace algunos años el uso de Internet estuvo en manos de los expertos, dada la cantidad de conceptos y comandos que las personas debían conocer para poder entrar al mundo cibernético [GIR, 2000]. Por mucho tiempo la gente había soñado con el concepto de un programa universal para poder tener acceso a cualquier tipo de información, que fuera sencillo e intuitivo de manejar, de tal forma que cualquiera pudiera utilizarla sin mayor complicación; además, sería ideal que este programa encontrara siempre la información que las personas desean a través de unas simples instrucciones. Pensando en esto, se comenzó a desarrollar sistemas que pudieran ser usados por personas con pocos conocimientos y experiencia en sistemas de cómputo. Estos sistemas incorporaron el uso del ratón (mouse) y pantallas gráficas, que minimizaron el uso de comandos [FER, 2002]. De esta manera nació el World Wide Web: un sistema creado a principios de los años 90 por Tim Berners-Lee [BER, 2001a] director del consorcio World Wide Web [W3C, 2005], que se ha convertido en un instrumento de uso cotidiano en nuestra sociedad, comparable a otros medios tan importantes como la radio, la televisión o el teléfono, a los que aventaja en muchos aspectos. La web es hoy un medio extraordinariamente flexible y económico para la comunicación, el comercio y los negocios, ocio y entretenimiento, acceso a información y servicios, difusión de cultura, etc. [CAS, 2003a].

Actualmente, la web es un espacio preparado para el intercambio de información diseñado para el consumo humano. Las páginas Web son creadas por personas para ser entendidas por personas. La información se publica en la web utilizando lenguajes que describen la forma en que ésta debe ser presentada al usuario por el navegador, pero no expresan nada sobre su significado y dificultan su manipulación [LOZ, 2003].

Otra carencia de la situación actual es que, con los estándares web del momento, no se puede diferenciar entre información personal, académica, comercial, etc. Es decir, cuando un buscador web realiza una consulta con algunas palabras clave, normalmente aparece información que no es útil por que en realidad no corresponde a lo que se está buscando y deja de presentarse información que a pesar de estar relacionada, no concuerda con las palabras clave utilizadas. Además no todas las páginas proporcionan igual cantidad de información, debido precisamente a que no existe un formato o convenio que nos diga qué contenido debemos añadir a las páginas web [SAU, 2002].

En los últimos años, importantes instituciones alrededor del mundo se han agrupado con el fin de diseñar modelos para transformar la red desde un espacio de información a un espacio de conocimiento. Tim Berners-Lee, uno de los inventores de la World Wide Web, defiende el desarrollo de la web con conocimiento, y organizaciones como Semantic Web se encargan de estandarizar lenguajes y herramientas para hacer efectiva la web semántica: una extensión de la web actual, en la cual, la información estará dotada de significados bien definidos, mejorando la forma de trabajo en cooperación entre humanos y computadores para solventar las limitaciones que actualmente posee la web [BER, 2001b]

Existe una aproximación al problema de la recuperación de información que, sin perseguir altas tasas de exhaustividad<sup>1</sup>, pretende obtener resultados mucho más precisos, basándose para ello en el uso de agentes inteligentes<sup>2</sup>. Aunque actualmente estos agentes no se diseñan para "comprender" la información que reside en la web, precisamente por que es prácticamente imposible conocer la representación de los datos ubicados en cada una de las páginas publicadas [LOZ, 2003].

La idea es que los datos puedan ser utilizados y comprendidos por los computadores sin necesidad de supervisión humana [LOZ, 2003], de forma que los agentes puedan ser diseñados para tratar la información situada en las páginas de la red, de manera semiautomática. Con ello, se mejorará la búsqueda de información y se potenciará el desarrollo de aplicaciones de comercio electrónico, entre otras, ya que las anotaciones de

---

<sup>1</sup> Numero de documentos relevantes recuperados dividido entre el número de documentos totales relevantes de la base de datos [MAR, 2002].

<sup>2</sup> programas informáticos con una serie de características especiales en los que es posible delegar responsabilidades [IWO, 2003].

información seguirán un esquema común al igual que los buscadores web pudiendo intercambiar datos siguiendo estos esquemas comunes consensuados [LOZ, 2003].

Estos agentes no sólo encontrarán la información de forma precisa, sino que podrán realizar inferencias automáticamente buscando información relacionada con la que se encuentra situada en las páginas y con los requerimientos de la consulta indicada por el usuario [LOZ, 2003].

Para que esto pueda llevarse a cabo, se necesita que el conocimiento de la web esté representado de forma que sea legible por los computadores, esté consensuado, y sea reutilizable. Las ontologías proporcionan la vía para representar este conocimiento a través de conceptos, relaciones, funciones, instancias y axiomas y los agentes Web podrán ser diseñados para tratar este conocimiento de manera semiautomática [LOZ, 2003].

Surge entonces la inquietud de plantear un Proyecto de Grado que aborde el campo de la semántica de la información en la Web y que motive la investigación al interior del Departamento de Sistemas e impulse el desarrollo de proyectos futuros que complementen lo obtenido en el mismo. Este proyecto está centrado en el desarrollo de un sistema de optimización de búsquedas que con base en el modelado de un dominio específico, filtrará el conjunto de páginas entregadas por motores de búsqueda convencionales, asegurando un resultado más preciso y acorde con los requerimientos de búsqueda del usuario.

El aporte de este trabajo de grado es la aplicación de técnicas emergentes en los campos de Sistemas Multiagente, Web Semántica y representación del conocimiento, para desarrollar un sistema de optimización de búsquedas en el dominio Mampostería Estructural<sup>3</sup> [HER,2003] y un prototipo de portal Web que presente al usuario de manera personalizada la información modelada; entregándole resultados concretos que le eviten pérdidas de tiempo en dispendiosos procesos de clasificación y análisis de la gran cantidad de páginas que usualmente entregan los buscadores tradicionales, resultados que si bien algunas veces logran satisfacer los requerimientos de búsqueda, en la mayoría de los casos no poseen la información que el usuario requiere.

---

<sup>3</sup>Elaboración de estructuras mediante la disposición ordenada de unidades de mampostería, cuyas dimensiones son pequeñas comparadas con las del elemento que se va a construir.

Cabe resaltar que para este proyecto se eligió la temática Mampostería Estructural dada la importancia que en los últimos años ha adquirido para el sector de la construcción en Ingeniería Civil al aportar no sólo las funciones estructurales sino también inmensas posibilidades y ventajas en lo que respecta a acabados; Sin embargo, dado que sus materiales poseen características particulares que los diferencian de otros, es necesario que quien vaya a manejarla tenga claridad, sobre sus características y los procesos que se deben seguir, con el fin de evitar que se incurra en imprecisiones o se omitan detalles que puedan poner en peligro la estructura de mampostería, demeritando su calidad y generando perjuicios para el usuario. De ahí la importancia para el sector constructivo de contar con una herramienta que permita optimizar el proceso de búsqueda de información en esta área y brinde información precisa que le permita al constructor conocer las ventajas constructivas, económicas y estéticas de la mampostería Estructural.

Se trata de un proyecto que integra dos áreas del conocimiento a través de la aplicación de conceptos de Ingeniería de Sistemas relacionados con la Arquitectura de la Información y de Ingeniería Civil pertenecientes al sistema constructivo denominado Mampostería Estructural.

El contenido de este documento presentará una descripción del trabajo realizado durante el desarrollo del proyecto a través de nueve capítulos, los cuales se exponen brevemente a continuación:

- **Capítulo 1: ESTADO DEL ARTE: DESARROLLO Y EVOLUCION DE LA WEB ACTUAL HACIA LA WEB SEMANTICA**  
Contiene una breve descripción del desarrollo y evolución de la Web actual hacia la Web Semántica, describiendo el importante papel que juegan los Sistemas de Representación del Conocimiento, Sistemas de Recuperación de Información y Agentes Software en esta evolución.
- **Capítulo 2: MODELO DEL DOMINIO.** Este capítulo contiene una descripción del proceso de modelado del dominio Mampostería Estructural, el estudio de metodologías, herramientas y lenguajes, así como la selección de cada uno de ellos para la implementación de la ontología.
- **Capítulo 3: SISTEMA MULTIAGENTE (Multi-Agent System MAS).** Este capítulo contiene una descripción del proceso de desarrollo del sistema multiagente.

- Capitulo 4: HERRAMIENTAS TECNOLÓGICAS EMPLEADAS PARA LA IMPLEMENTACION DEL SISTEMA. MULTIAGENTE. En este capítulo se describen todas las herramientas tecnológicas empleadas durante el desarrollo del sistema multiagente.
- Capitulo 5: PROTOTIPO DE PORTAL WEB. En este capítulo se describen las características del prototipo de portal web, por medio del cual se probará el sistema de filtrado de información.
- Capitulo 6: PRUEBAS Y CONCLUSIONES. En este capítulo se presenta el diseño de las pruebas realizadas con el sistema multiagente desarrollado y las conclusiones de este trabajo de grado.
- Capitulo 7: REFERENCIAS.
- Capitulo 8: ANEXOS. En este estudio se realiza una revisión teórica de distintos sistemas de búsqueda de información en Internet (motores, índices, metabuscadores y sistemas de filtrado), un estudio de metodologías de modelado, herramientas y plataformas para el desarrollo de aplicaciones basadas en agentes y un estudio de metodologías de desarrollo e implementación de ontologías.



# 1 ESTADO DEL ARTE: EVOLUCIÓN DE LA WEB ACTUAL HACIA LA WEB SEMÁNTICA.

## 1.1 WEB SEMÁNTICA

Actualmente la información presente en la web se encuentra estructurada mediante lenguajes de etiquetado que únicamente describen la forma en que dicha información debe ser presentada al usuario por el navegador (colores, fuentes, etc.), pero no expresan nada sobre su significado, es decir, su semántica [LOZ, 2003]. El objetivo de la web semántica es que toda esta información sea comprensible no sólo por humanos, sino también por computadoras. Para conseguir esto se deberá codificar la semántica de los documentos web mediante lenguajes de metadatos y ontologías [MON, 2003].

La web semántica es entendida como una extensión de la actual web que permitirá encontrar, compartir y combinar la información más fácilmente y como una red en la que los computadores no sólo serán capaces de presentar la información contenida en las páginas web, como hacen ahora, sino que además podrán "entender" dicha información [BER, 2001b].

En la última década ha existido una preocupación constante por que las páginas sean comprendidas por los usuarios, pero pocos se han preocupado por hacer que estas páginas sean entendidas por el software. La web semántica propone superar las limitaciones actuales de Internet mediante la introducción de descripciones explícitas del significado, la estructura interna y la estructura global de los contenidos y servicios disponibles en la web [BER, 2001b].

Esta semántica permitirá que agentes inteligentes interpreten el significado de los documentos, con lo que podrían asistir a usuarios en tareas como, por ejemplo, la recuperación de información o la gestión de su agenda.

En la práctica esto significa que las máquinas – computadores personales o cualquier otro dispositivo conectado a Internet– podrán realizar, casi sin necesidad de intervención humana, infinidad de tareas que simplificarán la vida del ser humano. "La web semántica no es una web aparte sino una extensión de la actual en la que la información tiene un significado bien definido, posibilitando que los computadores y las personas trabajen en

cooperación" [BER, 2001b]. Así, lo que hasta ahora había sido básicamente "un medio de documentos para personas" pasará a ser un sistema de datos e información que se podrán procesar automáticamente, gracias a la utilización de sistemas como las ontologías para representar la semántica de los datos presentes en la red.

## **1.2 ONTOLOGIAS EN WEB SEMANTICA**

Ontología es una antigua disciplina que en sentido filosófico, se define como un esquema específico de categorías que refleja una visión específica del mundo. Apartándose del aspecto filosófico, se define ontología como una teoría que explica cómo un individuo, grupo, lenguaje o ciencia entiende un determinado dominio [RED, 2004].

Desde el punto de vista informático las ontologías son artefactos que especifican un vocabulario relativo a un cierto dominio, definiendo conceptos y relaciones en el mismo, de forma compartida y consensuada. Esta conceptualización debe ser representada de una manera formal, legible y utilizable por los computadores, de ahí que las ontologías jueguen un papel clave en la interoperabilidad semántica entre sistemas de información, los cuales deberán comprometerse con el vocabulario que se maneje en las ontologías que comparte [ALC, 2004].

## **1.3 SISTEMAS MULTIAGENTE EN LA RECUPERACION DE INFORMACION EN LA WEB**

Uno de los fenómenos más importantes de los últimos años en el campo de la Información es el desarrollo y espectacular crecimiento de Internet, especialmente de lo que conocemos como Web. El número de páginas crece exponencialmente y afecta a todos los ámbitos del conocimiento. En el terreno científico, por ejemplo, muchas de las fuentes de información tradicionales como revistas, actas de congresos, etc. se encuentran ya en la red. Actualmente es posible observar como la red favorece la aparición de nuevos tipos de fuentes de información. Incluso algunas de ellas ya sólo se encuentran allí, dejando abandonados los soportes tradicionales. Este hecho pone de relieve el problema de la recuperación de información en la red, y más específicamente en la web [BER, 2001].

En la actualidad existe una gran cantidad de definiciones para el término recuperación de Información y es un hecho bastante curioso, que un término tan empleado, presente cierta confusión a la hora de establecer una definición que lo sitúe dentro del campo de las

ciencias de la información [MAR, 2002]. El uso frecuente de este término, ha propiciado que el mismo, no se encuentre bien empleado y resulta muy común, que gran cantidad de autores no profundicen en cuanto a las diferencias existentes entre “Recuperación de Datos” y “Recuperación de Información”. Estos dos términos son presentados como sinónimos, llegando a olvidar, que se puede recuperar información sin procedimientos informáticos, aunque evidentemente, esto no es lo más común hoy en día, pero el frecuente y necesario empleo de una tecnología no debe sustituir el adecuado uso de los conceptos terminológicos [MAR, 2002].

Para empezar, la principal tarea de la recuperación de información es traer documentos relevantes desde un gran repositorio como respuesta a una consulta del usuario y ordenar estos documentos de acuerdo con su relevancia [PER, 2000]. También se afirma que la recuperación de información no es encontrar simples correspondencias a unos patrones de bits, es encontrar documentos verdaderamente relevantes [GRO, 1998].

Por último, intentando definir el término recuperación de información, se exponen las diferencias entre recuperación de información y recuperación de datos, destacando que los datos se pueden estructurar en tablas, árboles, etc., para recuperar exactamente lo que se quiere, pero el texto, no posee una estructura clara y por tanto no resultará fácil crearla [BAE, 1999]. El problema de la recuperación de información se puede definir como: “dada una necesidad de información (consulta+perfil del usuario+...) y un conjunto de documentos, ordenar los documentos de mas a menos relevantes para esa necesidad y presentar un subconjunto de aquellos de mayor relevancia” [BAE, 1999].

Desde hace algunos años, la tecnología de agentes, se ha convertido en una fuente de soluciones al problema de la recuperación de información en la web. En este momento se habla de la web semántica, sistemas de manejo de base de datos auto administrables, bioinformática, sistemas de recuperación de información, etc., en donde la tecnología basada en agentes se torna fundamental [GAR, 2002].

La recuperación de información en la web es solo una de las múltiples tareas que ha sido abordada mediante agentes, dado que un agente es un sistema de computador capaz de realizar acciones independientes en beneficio de su propietario o de un usuario, en este caso particular, su finalidad primordial es recuperar páginas relevantes para unas necesidades informativas determinadas, siendo la formulación de tales necesidades parte de las especificaciones iniciales proporcionadas a los agentes; estos recopilan automáticamente la información que satisface esas especificaciones de forma muy

parecida a como lo hacen manualmente las personas, sin requerir de su presencia [BERR, 2001].

Aunque no es fácil conceptualizar los agentes software, existen una serie de atributos que los caracterizan y que aunque no todos los agentes poseen a la vez todas las características, estas son sumamente útiles cuando se enfrenta el problema de la recuperación de información en la web. Los agentes software son entidades autónomas y pueden trabajar cooperativamente para lograr un objetivo común, sin requerir constantemente la intervención humana para realizar su trabajo. Son reactivos, porque tienen algún medio de percibir su entorno y actuar sobre él y pro-activos, ya que para lograr un objetivo pueden elaborar un plan. Un grupo de dos o más agentes trabajando en colaboración es llamado sistema multiagente, este grupo de agentes puede distribuirse y comunicarse entre sí y la característica de sociabilidad se convierte en una herramienta fundamental para solucionar problemas complejos como la recuperación de información, imitando en ocasiones modelos de interacción con características humanas (antropomórficos) [PAV, 2004].

Puesto que la propia exploración de la web, manual o automática, requiere grandes cantidades de tiempo, no es esperable una respuesta inmediata, ni probablemente con la agilidad suficiente para plantear una dinámica especialmente interactiva con el usuario. Antes bien y muy en la línea de lo que se entiende por agentes inteligentes, el usuario delega en los agentes después de facilitarles algunas instrucciones, los agentes hacen su trabajo de forma autónoma y tomándose su tiempo, en espera de que en un plazo razonable, entreguen un resultado, esto es, las páginas Web encontradas útiles para satisfacer las necesidades de información expresadas por el usuario [BERR, 2001]. La mejora de los mecanismos de trabajo actuales y su aplicación en la recuperación de información al integrarse con tecnologías de agentes y multiagente, aseguran el logro de aplicaciones muy precisas en sistemas de usuario, a pesar del problema de los relativamente largos tiempos de espera, que mejoran con la característica de cooperación propia de los sistemas multiagente [BERR, 2001].

## 1.4 MODELO MATEMATICO PARA RECUPERACION DE INFORMACION

### 1.4.1 Modelo Vectorial

El modelo vectorial fue definido por Salton [SAL, 1968] hace ya bastantes años, y es ampliamente usado en operaciones de recuperación de información, así como también en operaciones de categorización automática, filtrado de información, etc. En el modelo vectorial se intenta recoger la relación de cada documento  $D_i$  de una colección de  $N$  documentos, con el conjunto de las  $m$  características de la colección. Formalmente un documento puede considerarse como un vector que expresa la relación del documento con cada una de esas características.

$$D_i \rightarrow \vec{d} = (C_{i1}, C_{i2}, \dots, C_{im}) \quad (1)$$

Es decir, ese vector identifica en qué grado el documento  $D_i$  satisface cada una de las  $m$  características. En ese vector,  $C_{ik}$  es un valor numérico que expresa en qué grado el documento  $D_i$  posee la característica  $k$ . El concepto “característica” suele concretarse en la ocurrencia de determinadas palabras o términos en el documento, aunque nada impide tomar en consideración otros aspectos. Si se consideran los términos como características definitorias del documento, el proceso que debe seguir el sistema pasa primero por seleccionar aquellos términos útiles que permitan discriminar unos documentos de otros. En este punto debemos señalar que no todas las palabras contribuyen con la misma importancia en la caracterización del documento. Desde el punto de vista de la recuperación de información existen palabras casi vacías de contenido semántico, como los artículos, preposiciones o conjunciones, que son poco útiles en el proceso. Pero también son poco importantes aquellas palabras que por su frecuencia de aparición en toda la colección de documentos pierden su poder de discriminación. En recuperación de información todas ellas forman parte del conjunto de palabras vacías (stops words en inglés), que se eliminan en el proceso de indexación. Una vez seleccionado el conjunto de términos caracterizadores de la colección de documentos, es necesario obtener el valor de cada elemento del vector del documento. El caso más simple es utilizar una aproximación binaria, de forma que si en el documento  $D_i$  aparece el término  $k$ , el valor  $C_{ik}$  sería 1, y en caso contrario sería 0. No obstante, una palabra puede aparecer más de una vez en el mismo documento, y además, algunas palabras pueden considerarse con más peso, esto es, más significativas que otras, de forma que el valor numérico de cada

uno de los componentes del vector obedece normalmente a cálculos más sofisticados que la simple asignación binaria. De otro lado, también es importante normalizar los vectores para no privilegiar documentos largos frente a otros documentos menos extensos (ecuación 2).

$$\vec{d}_i = \frac{1}{\sqrt{\sum_{j=1}^m w_{ij}^2}} (w_{i1}, w_{i2}, w_{i3}, \dots, w_{im}) \quad (2)$$

Se han propuesto diversos métodos para calcular el peso de cada término en el vector documento [SAL, 1983], [SAL, 1988], [HAR, 1992a], pero en general, para estimarlos se parte de dos ideas en cierto sentido contrapuestas: si un término aparece mucho en un documento, es importante para caracterizar ese documento. Pero si aparece en muchos documentos de la colección, no es beneficioso para distinguir un documento de los demás, dado su escaso poder discriminatorio, resultando poco útil para la recuperación.

Para determinar la capacidad de representación de un término para un documento dado se computa el número de veces que aparece en dicho documento, obteniéndose la (tf) frecuencia del término en el documento.

Por otra parte, si la frecuencia de un término en toda la colección de documentos es extremadamente alta, se opta por eliminarlo del conjunto de términos de la colección (pertenece al conjunto de palabras vacías). Podría decirse que la capacidad de recuperación de un término es inversamente proporcional a su frecuencia en la colección de documentos. Esto es lo que se conoce como idf (inverse document frequency). Así, para calcular el peso de cada elemento del vector que representa al documento se tiene en cuenta la frecuencia inversa del término en la colección, combinándola de alguna forma con la frecuencia del término dentro de cada documento. Normalmente se utiliza para ello el producto simple el cual viene dado por la ecuación 3 [HAR, 1992].

$$w_{ij} = tf_i \cdot idf_j \quad (3)$$

Algunos autores experimentaron con más de 200 sistemas de cálculo de pesos [SAL, ], pero uno de los más utilizados viene dado por la ecuación 4, que expresa el peso del término j en el documento i,

$$w_{ij} = tf_{ij} \cdot \log \frac{N}{df_j} \quad (4),$$

donde  $df_j$  es el número de documentos en que aparece el término  $j$ , y  $N$  el número de documentos de la colección.

El proceso realizado para los documentos también puede aplicarse a las consultas. Efectivamente, una consulta,  $Q$ , realizada en lenguaje natural está formada por términos y por tanto, puede verse como un documento más, seguramente bastante breve, aunque no siempre. Así pues, el mecanismo de obtención de pesos también se aplica a las consultas, para de esta manera poder disponer de representaciones homogéneas de consultas y documentos, que posibiliten obtener el grado de similitud entre ambas representaciones. El vector representante de la consulta está formado por un vector de igual número de elementos que los vectores de los documentos. Cada elemento de ese vector expresa el grado en que cada uno de los términos de la colección representa las necesidades informativas de la persona que hace la consulta (ecuación 5).

$$Q \rightarrow \vec{q} = \frac{1}{\sqrt{\sum_{j=1}^m p_j^2}} (p_1, p_2, \dots, p_m) \quad (5)$$

La resolución de la consulta consiste en un proceso de establecer el grado de semejanza entre el vector consulta y el vector de cada uno de los documentos. Para una consulta determinada, cada documento arrojará un grado de similitud determinado; aquéllos cuyo grado de similitud sea más elevado se ajustarán mejor a las necesidades expresadas en la consulta, desde el punto de vista del sistema de recuperación de información. No obstante, es el usuario el que debe decidir la relevancia de los documentos recuperados, siendo ésta una característica totalmente subjetiva del mismo.

El modo más simple de calcular la similitud entre una consulta y un documento, utilizando el modelo vectorial, es realizar el producto escalar de los vectores que los representan (ecuación 6). En esa ecuación se incluye la normalización de los vectores, a fin de obviar distorsiones producidas por los diferentes tamaños de los documentos. El índice de similitud más utilizado es el valor de la distancia entre ambos vectores. Para una consulta  $Q$ , el índice de similitud con un documento  $D_i$  es:

$$Dis\ tan\ cia(Q, D_i) = \frac{\sum_{j=1}^m p_j d_{ij}}{\sqrt{\sum_{j=1}^m p_j^2 \cdot \sum_{j=1}^m d_{ij}^2}} \quad (5)$$

Hay otros métodos propuestos para calcular la similitud [SAL, 1983]. Los resultados de la computación del índice de similitud entre la consulta y todos los documentos permiten ordenar los resultados en orden decreciente. De esta manera se le ofrecen al usuario primero los documentos que el sistema de recuperación considera más similares con la consulta, y que pueden coincidir, o no, con lo esperado por el usuario. La relevancia es la medida subjetiva que el usuario tiene para determinar si los resultados, y en qué grado, son adecuados a sus necesidades informativas.

## 1.5 TRABAJOS RELACIONADOS

En la actualidad existe un gran número de personas y entidades que trabajan en el desarrollo de proyectos que involucran las temáticas nombradas anteriormente en este documento. Para empezar, la web semántica se ha convertido en un área de investigación importante en varios centros de investigación de todo el mundo, entre ellos el MIT [MIT, 2003], la Universidad de Stanford con sus grupos Onto-Agents [DEC, 2002a], Scalable Knowledge Composition [DEC, 2002b] y Protégé [VEN, 2004], por citar tan sólo algunos de los grupos más fuertes. En pocos años se ha consolidado una comunidad investigadora considerable, de cuyo reflejo cabe destacar el gran congreso internacional que se celebra anualmente International Semantic Web Conference (ISWC) [ISW, 2004], y revistas como el Journal of Web Semantics [GOB, 2005], o el área The Semantic Web de Electronic Transactions on Artificial Intelligence (ETAI) [ETA, 2004]. Es muy importante además destacar el apoyo y el importante papel del W3C [W3C, 2001] en el proyecto de la Web semántica, con la creación de grandes y muy activos grupos de trabajo para el desarrollo de esta área, y muy en especial liderando el esfuerzo de estandarización de lenguajes y tecnologías específicas necesarias para el desarrollo de la web semántica [CAS, 2003a].

Por otra parte, la investigación en recuperación de información es un área que conoce en la actualidad un desarrollo sin precedentes. Uno de sus principales atractivos reside en su carácter esencialmente multidisciplinar, participando de muy diversos ámbitos del



conocimiento: Ciencias de la Documentación, Informática, Matemáticas, Lingüística y otros [FIG, 2003]. El motor de recuperación de información KARPANTA es solo un ejemplo de los numerosos proyectos que se han desarrollado en el área de recuperación de Información. KARPANTA es un motor de recuperación extremadamente flexible, que implementa un gran número de algoritmos diferentes, y que aísla el proceso de indización automática y resolución de consultas de las fases de análisis léxico y visualización [FIG, 2003].

La teoría de agentes ha entrado a jugar un papel importante en el desarrollo de herramientas para manejar el problema de sobrecarga y recuperación de información en la web. Un tipo de herramienta de amplio uso son las de Diseminación Selectiva de Información, en las que se busca recuperar información de la web y que además esta información llegue al usuario de una manera selectiva [GAR, 2002]. El “Sistema de Diseminación Selectiva de Información en la web” desarrollado en la Universidad Autónoma de Bucaramanga, Colombia, es un ejemplo que vale la pena nombrar, entre otros, no solo por que involucra la tecnología de agentes en su desarrollo sino que además de plantear el uso de la metodología GAIA para el modelado; también propone la integración de GAIA y AURL (Agent-Unified Modeling Language) con el fin de obtener un diseño cercano a la implantación que le permitirá al desarrollador una mejor comprensión del sistema multiagente que se construye [GAR, 2002].

TodoBr [AKW, 2005] y TodoCl [BAE, 2000] son proyectos que persiguen objetivos similares a los del nuestro, tales como facilitar el acceso a Información relevante pero limitando el alcance, en este caso, a contextos geográficos determinados. TodoBr es una máquina de búsqueda, enteramente propietaria, que tiene características de relevancia y velocidad de búsqueda de resultados que la diferencian de cualquier otra máquina de búsqueda con foco en la Internet brasileña. TodoCl es una máquina de búsqueda en Internet que hace parte de una familia de buscadores que incluyen a TodoBr, para sitios en Brasil, y a TodoLatin, para localizar sitios a nivel regional. El buscador TodoCl, tiene como objetivo mejorar la calidad de las respuestas encontradas, entregando información más precisa en menos tiempo, cuando se necesita algo ubicado en la web chilena.

En Colombia, el proyecto TodoCo: buscador semántico colombiano planteado en la facultad de ingeniería electrónica de la Universidad del Cauca, surge como idea de “generar un buscador semántico que permita obtener resultados más coherentes y de más

rápido acceso al entorno colombiano. Inicialmente se propone trabajar con información estructurada en RDF acerca del turismo colombiano, y se espera posteriormente estructurar TodoCo como un portal colombiano que involucre otras temáticas y servicios a partir de la información existente” [COL, 2004].

Otra temática relacionada con lo propuesto en este proyecto de filtrado de información son los sistemas de recomendación colaborativos también denominados filtros colaborativos o sistemas de recomendación social, los cuales se definen, como aquellos sistemas en que las recomendaciones son hechas basándose solamente en los términos de similitud entre los usuarios, es decir, no se realiza ningún tipo de análisis sobre los objetos a recomendar sino que las recomendaciones se llevan a cabo por aquellos usuarios que han mostrado intereses y preferencias similares en el pasado [BAL, 1997], [PEM, 2000]. A diferencia de los filtros colaborativos, el sistema de filtrado planteado en este documento, analiza los objetos (documentos) a recomendar y no toma en cuenta la opinión de los usuarios, pero es una excelente alternativa y se podría plantear como un buen complemento para el sistema de filtrado de información puesto que además de analizar automáticamente los documentos a recomendar, tomaría la opinión de los usuarios para mejorar el proceso de filtrado. Un ejemplo de sistemas de recomendación colaborativos es el proyecto “A Model for a Collaborative Recommender System for Multimedia Learning Material”, el cual sugiere al estudiante material multimedia de aprendizaje basado en sus preferencias y en las características software y hardware disponibles por él [BAL, 2004] y propone una metodología para describir el material a través del uso de técnicas colaborativas definiendo un vector de características para cada recurso, el cual refleja la opinión de los usuarios que lo han visto anteriormente, dándole un mayor peso a la opinión dada por aquellos usuarios con intereses similares y prediciendo la opinión que nuevos usuarios podrían tener sobre el mismo. Adicionalmente a la opinión expresada por los usuarios, el sistema también considera las posibilidades que el estudiante tiene de desplegar cierto tipo de material de aprendizaje, considerando las características del software, hardware y conexión de que el estudiante dispone en el momento de realizar la búsqueda y de proponerle el objeto de aprendizaje, ya que estas características pueden ser tomados en cuenta por él para decidir si descarga o no el material. La importancia de este proyecto para el trabajo que se presenta en este documento, radica en que su objetivo es brindar al usuario una búsqueda más eficiente sobre la información que se encuentra en la red, además de la utilización de tecnologías de agentes, el mantenimiento de los intereses del usuario sobre diferentes dominios y la caracterización de los distintos recursos web. Sin

embargo, dichos proyectos a diferencia del propuesto en este documento, no utilizan sistemas estandarizados de representación del conocimiento que puedan ser compartidos de manera eficiente en el entorno Web.

## 2 MODELO DEL DOMINIO

La web semántica es un área que ha surgido a partir de la unión entre la Ingeniería del Conocimiento y las tecnologías web y propone la aplicación de nuevas técnicas basadas en la introducción de conocimiento semántico explícito que describa y estructure la información y servicios disponibles, de manera que puedan ser fácilmente utilizados, compartidos y procesados tanto por seres humanos como de forma automática [RAI, 2004]. Uno de los factores más importantes para la realización de esta visión es el concepto de ontología como herramienta clave para alcanzar un entendimiento entre las partes (usuarios, desarrolladores, programas) que participan de este conocimiento común [SAI, 2004], [RAI, 2004].

Actualmente, la definición de ontología más ampliamente citada es la de Gruber: “Una ontología es una especificación formal y explícita de una conceptualización compartida. “Conceptualización” se refiere a un modelo abstracto del fenómeno en el mundo, mediante la identificación de conceptos relevantes de ese fenómeno. “Explícito” significa que el tipo de conceptos usados, y las restricciones en su uso están explícitamente definidos. “Formal” se refiere al hecho de que la ontología debe ser legible por la máquina. “Compartida” refleja que la ontología debe capturar conocimiento consensuado y aceptado por las comunidades” [GRU, 1993a].

En términos prácticos, una ontología es una descripción formal y explícita de conceptos en un dominio de discurso que usualmente se concibe como un conjunto de clases o conceptos, propiedades de cada concepto que describen sus características, atributos y restricciones, así como las relaciones existentes entre ellos [DIN, 2004].

Las ontologías tienen los siguientes componentes que servirán para modelar el conocimiento de algún dominio [GRU, 1993b]:

- **Conceptos:** son las ideas básicas que se intentan formalizar. Los conceptos pueden ser clases de objetos, métodos, planes, estrategias, procesos de razonamiento, etc.
- **Relaciones:** representan la interacción y enlace entre los conceptos del dominio. Suelen formar la taxonomía del dominio. Por ejemplo: subclase-de, parte-de, parte-exhaustiva-de, conectado-a, etc.
- **Funciones:** son un tipo concreto de relación donde se identifica un elemento mediante el cálculo de una función que considera varios elementos de la ontología. Por ejemplo, pueden aparecer funciones como categorizar-clase, asignar-fecha, etc.
- **Instancias:** se utilizan para representar objetos determinados de un concepto.
- **Axiomas:** son teoremas que se declaran sobre relaciones que deben cumplir los elementos de la ontología. Por ejemplo: “Si A y B son de la clase C, entonces A no es subclase de B”, “Para todo A que cumpla la condición C1, A es B”, etc [GRU, 1993b].

## **2.1 CRITERIOS DE DISEÑO DE ONTOLOGIAS**

Es ampliamente reconocido que la construcción de un modelo de dominio u ontología, es un paso importante en el desarrollo de sistemas basados en el conocimiento dadas las facilidades que proporcionan para compartir y reutilizar el conocimiento, así como para su recopilación, verificación y mantenimiento. No obstante, el proceso de construcción de dichos modelos es aún más un arte que un proceso de ingeniería [JON, 1997]

Existe un conjunto de principios que deben observarse durante el proceso de construcción de una ontología para evaluarla y verificar si la ontología obtenida está completa.

### **2.1.1 Principios de Diseño**

- **Claridad y Objetividad:** Una ontología debe efectivamente comunicar el significado deseado del término definido. Esta definición debe ser independiente del contexto social o computacional y además debe ser documentado en lenguaje natural.
- **Compleitud:** Una definición expresada en términos de condición necesaria y suficiente es preferible sobre una definición parcial (definida solo sobre condición necesaria o suficiente)

- Coherencia: Para asegurar (permitir) que las inferencias derivadas de esta sean consistentes con las definiciones.
- Extensibilidad: Una ontología debe ser diseñada para anticipar los usos que pueda tener al tratarse de un vocabulario compartido. Es decir, las especializaciones o generalizaciones se deben poder incluir en la ontología de tal forma que no se requiera una revisión de las definiciones preexistentes.
- Principio de distinción ontológica: Las clases en una ontología deben ser disjuntas.
- Diversificación: de las jerarquías incluidas para aumentar la potencia de los mecanismos de herencia múltiple.
- Modularidad: para disminuir el acoplamiento entre los módulos.
- Estandarización de nombres: siempre que sea posible.
- Minimización de la distancia semántica entre conceptos emparentados: Conceptos similares estarán agrupados usando las mismas primitivas [GRU, 1993a].

En cada una de las metodologías presentadas (ver anexos), se identificaron dos etapas generales; en la primera de ellas, se genera una descripción informal de la ontología y en la segunda etapa se formaliza la descripción mediante la implementación en un lenguaje determinado.

Existe la expectativa de construir repositorios de ontologías, que permitan su reutilización y que puedan ser compartidas. Con respecto a esto, se deben tener en cuenta algunos criterios para seleccionar posibles ontologías existentes, así como para su desarrollo pensando en que posteriormente pueda ser compartida, reutilizada y extendida.

Para el desarrollo de este proyecto, se requiere la aplicación estricta e iterativa y de una metodología precisa que garantice la obtención de un producto de alta calidad, que cumpla con los objetivos planteados. Con base en el estudio de varias metodologías propuestas por diferentes autores, se propone entonces un proceso a seguir que involucra apartes de las metodologías estudiadas previamente y que en conjunto constituirán el modelo o la metodología de desarrollo de una ontología en el dominio mampostería estructural para el proyecto “Sistema de optimización de búsquedas en el dominio Mampostería Estructural utilizando técnicas de Web Semántica y tecnologías Multiagente”.

## **2.2 METODOLOGIA ADAPTADA**

La metodología propuesta para el desarrollo de la ontología que representa el dominio mampostería estructural, fue creada con base en el estudio de metodologías propuestas para el desarrollo de otros proyectos y se divide en seis etapas:

### **2.2.1 Especificación**

El propósito de esta etapa, es identificar el propósito de la ontología, incluyendo las pretensiones de los usuarios, escenarios de uso, grado de formalidad requerida. etc., y el alcance de la ontología incluyendo un grupo de términos para ser representados, características y granularidad requerida [NOY, 2000]

Inicialmente se plantean algunas preguntas básicas que marcaran el inicio del proceso:

- ¿Cual es el dominio que cubrirá la ontología?
- ¿Para que será usada la ontología?
- ¿Quién usará y mantendrá la ontología?

Las respuestas a estas preguntas podrían cambiar en el transcurso del proceso de diseño de la ontología, pero, la ventaja es que en algún momento ellas proveerán información más precisa y relevante a los desarrolladores para determinar y limitar el alcance del modelo.

El resultado de esta etapa es un documento de especificación de requerimientos de la ontología DERO [GON, 2003], el cual incluye el dominio que cubre la ontología, el uso que se le va a dar y quienes son los usuarios finales de la misma.

### **2.2.2 Adquisición del conocimiento**

El propósito de esta etapa es recopilar la información necesaria sobre el dominio seleccionado, desde las diferentes fuentes disponibles para tal fin: Internet, manuales, códigos, documentación bibliográfica, seminarios y cursos en los que el grupo de trabajo ha participado, además de la asesoría de docentes e ingenieros constructores expertos en el tema.

Existen varias técnicas que se pueden aplicar durante esta etapa y es preciso aclarar que es necesario aplicar todas o gran parte de ellas ya que las distintas fuentes, en especial

los expertos, poseen diferentes tipos de conocimiento y por lo tanto se deben aplicar distintas técnicas que permitan acceder a cada uno de los tipos de conocimiento que cada experto posee [MOR, 1999].

#### 2.2.2.1 Obtención del conocimiento a partir del experto

En esta parte se aplican técnicas basadas en protocolos, que consisten en entrevistas con el experto, informes y técnicas basadas en observación [MOR, 1999].

Se lleva a cabo una entrevista inicial con el experto, la cual es recomendable grabar, con el objetivo de determinar el ámbito y propósito del conocimiento, adquirir y comprender la terminología del dominio. Se transcribe la entrevista inicial y se analiza el resultado.

Posteriormente, se realizan una serie de entrevistas con el experto, que permiten exponer los puntos de vista de los integrantes del grupo respecto de la temática a modelar, intercambiar opiniones y complementar el proceso a través de las demás fuentes de información a las que se tiene acceso, con el fin de ir refinando el conocimiento adquirido con el experto.

El resultado de esta etapa es una serie de documentos en los cuales se transcriben las entrevistas que se han sostenido con el experto en el área, las correcciones y modificaciones a que haya lugar durante el proceso.

Cabe aclarar que la adquisición del conocimiento es un proceso que se lleva a cabo durante todo el ciclo de vida de la ontología y más aún, durante todo el ciclo de vida del proyecto [MOR, 1999].

### **2.2.3 Conceptualización**

Se definen los términos como conceptos, instancias, verbos relaciones o propiedades más relevantes dentro del dominio Mampostería Estructural para lo cual, se han definido los siguientes pasos:

#### 2.2.3.1 Declaración de términos importantes en la ontología

Una vez terminada la etapa de entrevistas, se procede a la aplicación de técnicas de análisis de protocolos, que utilizan las transcripciones de entrevistas o información textual de distintos tipos para identificar distintos tipos de conocimiento (objetivos, relaciones, atributos). Se crea una jerarquía con los conceptos obtenidos aplicando los pasos citados a continuación, para comprender el tipo de conocimiento de este dominio y se usa este resultado para generar preguntas que serán resueltas con el experto; de forma iterativa se analizan los resultados arrojados por las entrevistas se utilizarán para refinar la jerarquía generada.

#### 2.2.3.2 Definición de clases y jerarquías de clases

Para definir las clases y la jerarquía de clases se aplica el proceso de desarrollo Top-down, el cual inicia con la definición de los conceptos más generales en el dominio y a continuación la especialización de esos conceptos [NOY, 2000].

#### 2.2.3.3 Definición de propiedades

Una vez definidas las clases, se describe su estructura interna definiendo sus propiedades conocidas como slots, debido a que las clases por si solas no proveen suficiente información para responder las preguntas iniciales.

En la etapa anterior se determinaron las clases y es posible que muchos de esos términos no queden etiquetados como clases, pero probablemente serán propiedades de esas clases. En general hay varios tipos de propiedades en una ontología, entre las cuales se podrían nombrar las siguientes:

- Propiedades intrínsecas: son aquellas propiedades que solamente las poseen unas cuantas clases.
- Propiedades extrínsecas: son aquellas propiedades generales para todas o gran cantidad de clases.
- Partes: Si el objeto esta compuesto por otros objetos, esta composición será representada por medio de slots.
- Relaciones con otras clases: slots que muestran las relaciones entre miembros individuales de la clase y otros ítems

#### 2.2.3.4 Definición de restricciones (facets) de los atributos



Los atributos pueden tener diferentes restricciones que describen el tipo de valor, entre las más comunes se encuentran los valores permitidos y la cardinalidad.

- Cardinalidad: Define que cantidad de valores puede tomar el atributo. La cardinalidad puede estipularse como simple (1), o múltiple (N). Esta también puede ser mínima o máxima.
- Tipo de valores: son todos aquellos tipos de valores que pueden tomar los atributos: String: es el tipo de valor más simple, Number: (number o float) describe atributos con valores numéricos, Boolean: son simples banderas si-no, Enumerados: lista de valores específicos para un atributo, Instance-type: permite la definición de relaciones entre clases, los atributos con este tipo de valores deben también definir una lista de clases permitidas desde las cuales pueden venir las instancias [NOY, 2000].

#### 2.2.3.5 Definición de características especiales: Observaciones.

Dentro del proceso es importante tener en cuenta la definición de propiedades con características como las siguientes:

- Atributos inversos

La información por ejemplo de comprador y productor de una misma cosa no es necesario tenerla en ambos sentidos, sería redundante, pero es necesario al inicio del modelado tener a disposición toda la información posible, después cuando se refine el modelo, se verá claramente que esta información solo es necesaria en uno de los dos sentidos.

- Valores por defecto

Si un atributo particular es el mismo para la mayoría de las instancias de una clase, es posible definir ese valor como un valor por defecto para el atributo, entonces cuando una nueva instancia de una clase que contenga ese atributo sea creada se llenará automáticamente el slot con el valor por defecto [NOY, 2000].

El resultado del trabajo realizado hasta este momento es un documento que describe el modelo del dominio, expresado en lenguaje natural, el cual se refinará durante todo el ciclo de vida de la ontología. Por último se procede con la siguiente etapa de desarrollo, la cual se describe a continuación.

## 2.2.4 Implementación

Durante la etapa de construcción de la ontología, surgen algunas preguntas relacionadas con las herramientas a usar: ¿Cuáles herramientas pueden ser las más convenientes para implementar?, ¿Qué clase de herramientas me dan un mejor soporte para el proceso de desarrollo de la ontología?, ¿Cómo se almacenan las ontologías en la herramienta (en Bases de Datos, en XML [W3C, 1999]?, ¿Tiene la herramienta un motor de inferencia?, ¿Tiene la herramienta traductores desde y hacia los diferentes lenguajes de implementación de ontologías?, ¿Cómo pueden interoperar las aplicaciones con las herramientas de construcción de ontologías?, ¿Cómo pueden ser usadas en aplicaciones reales las ontologías desarrolladas? ¿Qué herramientas se pueden usar para consultar información de una ontología?

Para dar respuesta a los interrogantes, es necesario estudiar algunos lenguajes de representación y herramientas para la edición de ontologías, evaluarlos y determinar cuales de ellos son los más convenientes para el desarrollo del proyecto.

Adicionalmente existe una clasificación de las herramientas de acuerdo a las tareas para las que van a ser utilizadas, lo cual será de gran importancia en el momento de decidir la herramienta y el lenguaje adecuados para el proyecto. La clasificación es la siguiente:

- Herramientas de desarrollo de ontologías: Este grupo incluye herramientas, entornos y suites que pueden ser usadas para construir una ontología totalmente nueva (caso de interés), o para reutilizar ontologías existentes. Además de las funcionalidades comunes de edición y navegación en la ontología, estas herramientas usualmente incluyen documentación, exportación e importación desde diferentes formatos, vistas gráficas de las ontologías construidas, librerías de ontologías, motores de inferencia, etc.
- Herramientas de mezclado e integración: Usadas para resolver el problema de mezclar o integrar diferentes ontologías sobre el mismo dominio, que surge cuando dos organizaciones trabajan juntas o cuando es necesario obtener una ontología de mejor calidad a partir de ontologías existentes en el mismo dominio.
- Herramientas de evaluación: Son herramientas de soporte que aseguran la calidad de las aplicaciones industriales basadas en ontologías.
- Herramientas de anotación basadas en ontologías: Son herramientas usadas para permitir anotaciones basadas en ontologías en páginas Web. La mayoría de ellas ya están integradas en los entornos de desarrollo de ontologías.

- Herramientas de almacenamiento y consulta: Son herramientas que han sido creadas para facilitar el uso y consulta sobre las ontologías.
- Herramientas de aprendizaje de ontologías: Son herramientas usadas para deducir, en forma semiautomática, ontologías a partir de textos en lenguaje natural [GOM, 2002].

El resultado final de esta etapa es la ontología implementada en un lenguaje determinado.

### **2.2.5 Evaluación**

En esta etapa del desarrollo se lleva a cabo una evaluación y una validación de la ontología desarrollada para verificación de inconsistencias redundancia y completitud de la misma. Este proceso se lleva a cabo inicialmente por medio de la colaboración del experto quien debe evaluar el modelo y emitir sus observaciones en procura de perfeccionar el modelo obtenido. Otra estrategia de evaluación es la integración de la ontología en el sistema completo para observar su comportamiento y su desempeño.

El resultado es un documento con las observaciones del experto y con una descripción de los resultados de la implantación de la ontología en el sistema global [NOY, 2000].

### **2.2.6 Documentación**

Esta es una actividad que se lleva a cabo en cada una de las etapas de la metodología y es la recopilación de los documentos resultados de las actividades que constituyen la metodología adaptada.

## **2.3 APLICACIÓN DE LA METODOLOGÍA ADAPTADA EN LA CONSTRUCCIÓN DE LA ONTOLOGÍA.**

### **2.3.1 Documento de Especificación de Requerimientos de la Ontología DERO**

#### **2.3.1.1 Descripción del dominio y el alcance de la ontología**

El dominio del conocimiento que se va a modelar corresponde a Mampostería Estructural, sistema constructivo que consiste en la elaboración de estructuras mediante la disposición ordenada de unidades de mampostería, cuyas dimensiones son pequeñas comparadas con las del elemento que se va a construir y cuyo peso y tamaño dependen del sistema de manejo que se vaya a emplear [GON, 2002].

En esta ontología inicial, se representan los conceptos más generales e importantes de la temática elegida, realizando una clasificación de los términos identificados como relevantes dentro del dominio (mampostería, elementos componentes y materiales) y la identificación de clases y subclases, con sus atributos más generales y las relaciones más importantes entre ellas, si las hay.

Por tratarse de una ontología básica inicial, no se profundiza en temas específicos del dominio como aquellos relacionados con el diseño, normas y especificaciones constructivas, dado que estos temas están por fuera de la finalidad del sistema en el cual va a ser usada la ontología. Los usuarios a quienes va dirigido este proyecto son los constructores que se desempeñan en la edificación de obras civiles en las cuales se ha adoptado como solución el sistema constructivo Mampostería Estructural, docentes en el área de la ingeniería civil y estudiantes.

#### 2.3.1.2 Áreas de aplicación potenciales

La ontología desarrollada puede ser usada para diversos fines, entre los que se pueden nombrar están los siguientes:

- Repositorios para la organización de conocimientos e información.
- Herramienta para la adquisición de información.
- Para permitir la reutilización del conocimiento ya existente, en la creación de nuevas aplicaciones.
- Como base para la construcción de lenguajes de representación del conocimiento.
- Para la indización, recuperación y divulgación de la información Web.
- Para ayudar en el trabajo de los agentes inteligentes de encontrar páginas especializadas en la Web: El agente debe ser capaz de identificar si una página dada que pertenece o no al dominio específico tratado.
- Para la creación de herramientas para mejorar la eficacia de búsquedas en la web: mediante el diseño de frames (framework) para la construcción, manipulación y navegación de ontologías, que permitan guiar al usuario en el proceso de búsqueda [LOZ, 2003].

La ontología va a ser implantada dentro de un sistema multiagente cuyo propósito es mejorar el resultado de búsquedas en la web dentro del área de Mampostería estructural, asegurando a los usuarios la obtención de información precisa sin tener que gastar mucho

tiempo en su análisis y clasificación; por esta razón, la representación del conocimiento estará fundamentada en una clasificación jerárquica de los conceptos involucrados en esta temática.

### **2.3.2 Conceptualización**

#### **2.3.2.1 Captura de entidades (conceptos) involucrados en el dominio**

La siguiente lista de términos es una aproximación al conjunto de conceptos más relevantes del dominio a modelar. Este grupo de términos, son objetos del dominio que inicialmente son considerados; pero no es estrictamente necesario que todos ellos hagan parte del modelo de dominio generado para el proyecto.

Acabados, Adherencia, Aditivos, Agregados, Aislamiento, Almacenamiento, Arena, Cal, Calidad, Cargas horizontales, Cargas verticales, Cemento, Cemento blanco, Colorantes, Conectores, Corta goteras, Chapas, Definición de mampostería, Densidad, Desventajas del sistema constructivo, Diseño arquitectónico, Diseño estructural, Durabilidad, Eflorescencias, Enchapes, Estucos, Fabricación, Fisuras, Fluidez, Herramientas, Hidrófugos, Humedad, Juntas flexibles, Juntas rígidas, Ladrillo, Lechadas, Limpieza, Mampostería, Mampostería estructural, Mampostería no estructural, Mano de obra, Mantenimiento, Medida, Modulación imperfecta, Modulación perfecta, Mortero de inyección, Mortero de pega, Mortero premezclado preparado en planta, Mortero premezclados preparados en obra, Normas, Pinturas, Recepción, Reconstrucciones, Recubrimientos, Refuerzo horizontal, Refuerzo vertical, Reparaciones, Resistencia, Resistencia a la compresión, Resistencia al fuego, Respiraderos, Tipo de junta, Tipos de bloque, Tolerancias, Transporte, Unidades de mampostería, Ventajas del sistema constructivo, Ventanas de inspección, Cara superior, Cara inferior, Pared, Extremo, Saliente, Tabique exterior, Tabique interior, Perforación vertical, Ranura para mortero.

#### **2.3.2.2 Clasificación jerárquica de conceptos en el dominio**

En la fase anterior se identificaron términos relevantes del dominio, posteriormente se consulto con el experto quien sugirió se descartaran términos de la lista, de la misma forma, se adicionaron nuevos términos y se obtuvo una clasificación jerárquica inicial, la cual se refinó con nueva información proveniente de diferentes fuentes y se constituyó la siguiente clasificación:

1. Tipos de Mampostería
  - 1.1 Mampostería
  - 1.2 Mampostería Arquitectónica
    - 1.2.1.1 Mampostería al Tope
    - 1.2.1.2 Mampostería Pegada
    - 1.2.1.3 Mampostería Expuesta
    - 1.2.1.4 Mampostería Cubierta
  - 1.3 Mampostería no Estructural
    - 1.3.1 Mampostería simple
    - 1.3.2 Mampostería Parcialmente Reforzada
  - 1.4 Mampostería Estructural
    - 1.4.1 Mampostería Reforzada
    - 1.4.2 Mampostería de cavidad reforzada
    - 1.4.3 Mampostería de muros diafragma
    - 1.4.4 Mampostería de muros confinados
2. Elementos Componentes
  - 2.1 Unidades de mampostería
    - 2.1.1 Bloque de arcilla
    - 2.1.2 Bloque de concreto
  - 2.2 Mortero
    - 2.2.1 Mortero de Pega
    - 2.2.2 Mortero de Inyección
  - 2.3 Refuerzo
    - 2.3.1 Refuerzo horizontal
    - 2.3.2 Refuerzo vertical
    - 2.3.3 Conectores
  - 2.4 Muros
  - 2.5 Acabados
    - 2.5.1 Acabados para interiores
      - 2.5.1.1 Revoque tradicional
      - 2.5.1.2 Revoque plástico
      - 2.5.1.3 Estucos tradicional
      - 2.5.1.4 Estucos plástico
      - 2.5.1.5 Pinturas
      - 2.5.1.6 Enchapes cerámicos
    - 2.5.2 Acabados para exteriores
      - 2.5.2.1 Pinturas
      - 2.5.2.2 Enchapes no cerámicos
      - 2.5.2.3 Estuco Plástico
      - 2.5.2.4 Estuco Tradicional
      - 2.5.2.5 Revoque Plástico
      - 2.5.2.6 Revoque Tradicional
      - 2.5.2.7 Recubrimientos texturizados
3. Materiales
  - 3.1 Cementos
  - 3.2 Agregados
  - 3.3 Aditivos
    - 3.3.1 Acelerantes
    - 3.3.2 Plastificantes
    - 3.3.3 Impermeabilizantes
    - 3.3.4 Retardadores de fraguado
  - 3.4 Acero de Refuerzo

### 2.3.2.3 Definición de propiedades y restricciones sobre las clases

Una vez realizada la definición de clases en la etapa anterior, se describe su estructura interna definiendo sus propiedades, restricciones y características especiales para cada una de ellas. El resultado es el modelo que se presenta a continuación:

La figura 1., ilustra la clasificación de los conceptos del dominio en trece categorías básicas, las cuales son usadas posteriormente en la integración del modelo con el sistema multiagente:

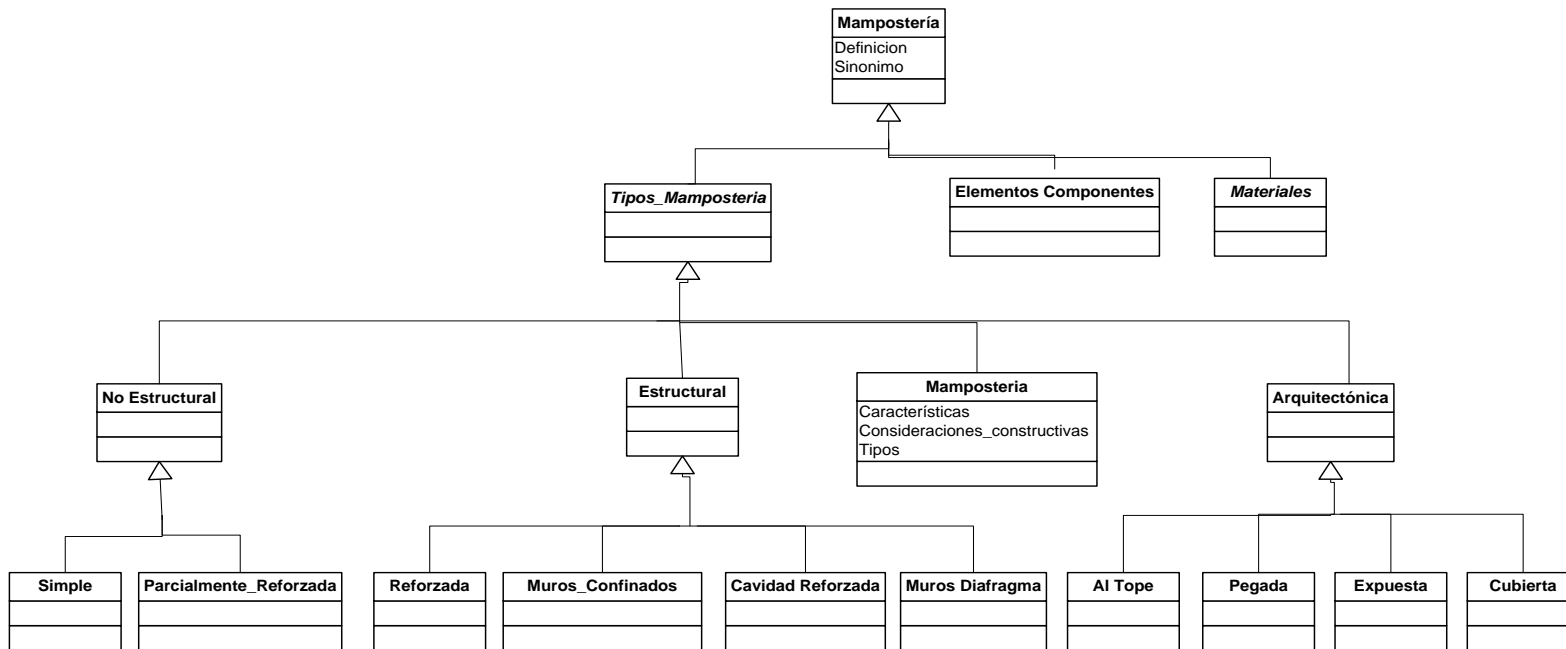


Figura 1. Clasificación de los conceptos del dominio.

La figura 2., ilustra la definición de propiedades y restricciones para las clases correspondientes a Elementos Componentes:

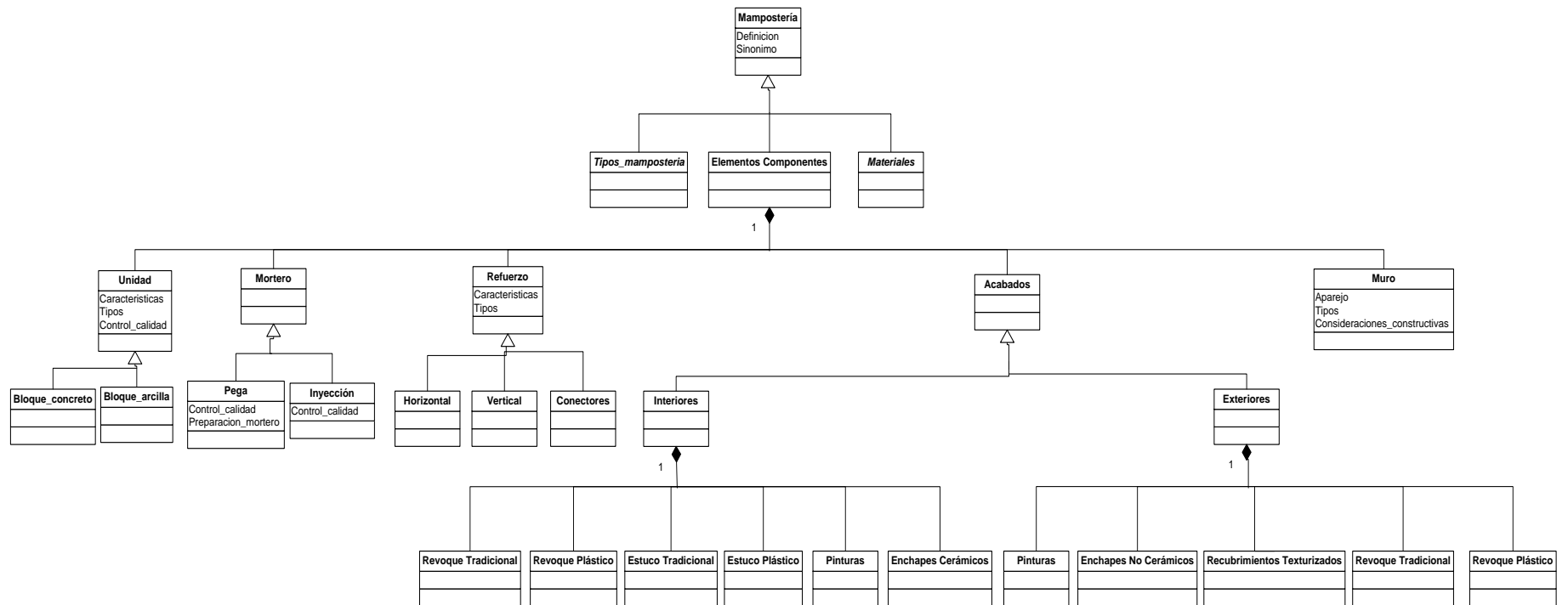


Figura 2. Definición de propiedades y restricciones para las clases correspondientes a Elementos Componentes



La figura 3., ilustra la definición de propiedades y restricciones para las clases correspondientes a Materiales.

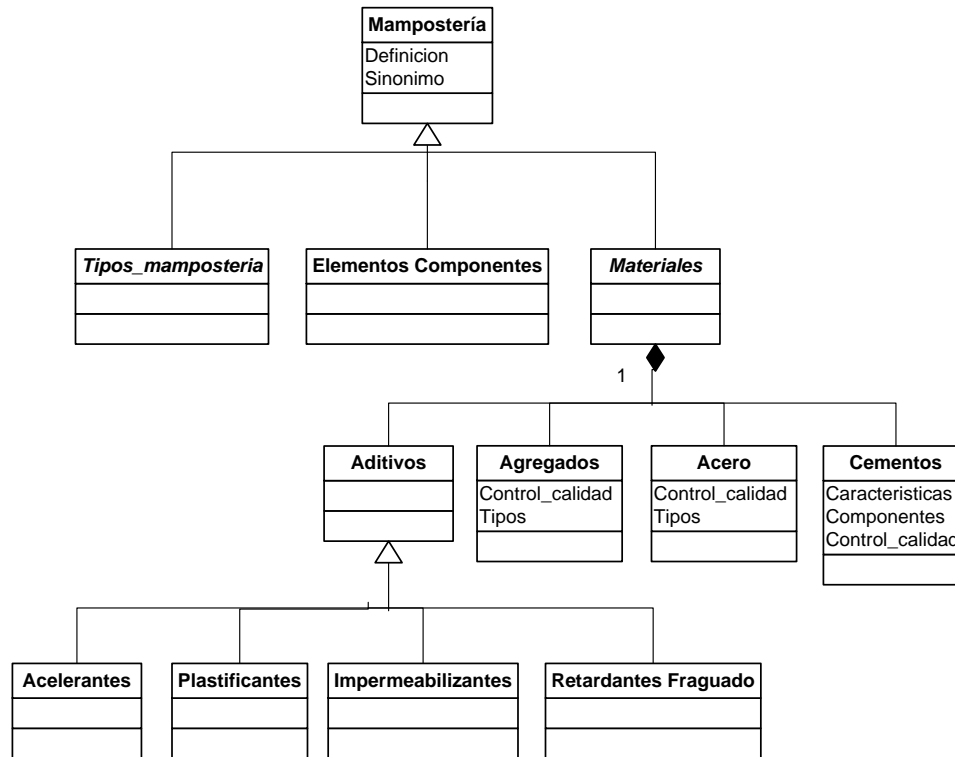


Figura 3. Definición de propiedades y restricciones para las clases correspondientes a Materiales.

### **2.3.3 Implementación**

Después de evaluar varias tecnologías y lenguajes para desarrollo de ontologías (ver anexos) el grupo de trabajo decidió que la herramienta adecuada para este proyecto es Protégé en su versión 3.1 [VEN, 2004] debido a que posee las características necesarias para la construcción de ontologías escalables y extensibles, facilidad de uso, soporte técnico y además es una herramienta de libre distribución que ha sido usada en gran cantidad de proyectos que involucran el desarrollo de ontologías para diferentes dominios de aplicación [VEN, 2004]. En cuanto al lenguaje de definición de ontologías se ha decidido que el más apropiado es OWL [W3C, 2004a] por ser este un lenguaje estándar de ontologías web, aunque, desde la perspectiva de éste proyecto, la selección de OWL u otro lenguaje no es un factor crítico, dado el nivel de abstracción que provee protege y su modelo de conocimiento basado en OKBC. Lenguajes anteriores han sido utilizados para desarrollar herramientas y ontologías destinadas a comunidades específicas (especialmente para ciencias y aplicaciones específicas de comercio electrónico) y no fueron definidos para ser compatibles con la arquitectura de la World Wide Web en general, y la web semántica en particular [HEN, 2004].

El resultado de esta etapa es un documento de implementación del modelo del dominio desarrollado con Protégé 3.1, representado en una ontología OWL, Esta ontología general se refina durante el desarrollo del proyecto y finalmente se integra al sistema global.

### **2.3.4 Documentación**

La documentación relacionada con la ontología OWL implementada se presenta en los anexos

### **2.3.5 Ontología de Consultas**

Adicionalmente, se desarrolló una ontología que extiende el modelo del dominio para representar las consultas que serán lanzadas al motor de búsqueda utilizado, para recuperar los documentos que serán analizados por el sistema multiagente.

Para la implementación de dicha ontología, se partió del modelo del dominio seleccionado, reemplazando la definición de propiedades por una única propiedad denominada consulta, la cual permite almacenar una serie de consultas previamente definidas para cada clase del dominio. La documentación de esta ontología se encuentra en la sección de anexos.

### 3 SISTEMA MULTIAGENTE (Multi-Agent System MAS)

En sus comienzos, el uso de la web estaba limitado al ámbito científico y académico, y estaba formada por un pequeño número de páginas que hacían manejable la poca cantidad de información que se encontraba en ella.

Internet se ha convertido sin duda alguna en la mayor fuente de información disponible. Actualmente alberga cerca de dos mil setecientos millones de páginas Web, que son usados por más de trescientos sesenta millones de usuarios en todo el mundo, adicionales a los millones de páginas más existentes en redes corporativas [OPI, 2004].

Toda esta información se expande por la web de forma caótica e incontrolada, ocasionando serias restricciones para el manejo, organización y recuperación de la información disponible en ella, lo que hace prácticamente imposible encontrar la información adecuada en el momento preciso.

La web fue diseñada como espacio de información, con el objetivo de que debería ser útil no solamente para la comunicación entre seres humanos, sino también para que los programas de computador pudieran manipularla y comprenderla.

Para el ser humano es fácil diferenciar entre información científica, de entretenimiento o de negocios gracias a su capacidad de interpretar la información que se le presenta, pero esta es una facilidad con la que no cuentan actualmente las aplicaciones software y que por su complejidad muy difícilmente se les podría otorgar.

Desarrollar aplicaciones que permitan aprovechar el potencial real de la web es demasiado complicado aún cuando toda esta información está disponible en ella, ya que es muy difícil dotar a las máquinas con la capacidad de una persona para comprender sus contenidos tal y como están codificados actualmente.

Los lenguajes utilizados para publicar la información en la web describen la forma en que ésta debe ser presentada al usuario, pero no expresan nada sobre su significado y dificultan su manipulación lo que impide explotar al máximo todo el potencial que la web nos ofrece. [MON, 2003].

La web semántica propone superar estas limitaciones de Internet mediante la introducción de descripciones explícitas del significado, la estructura interna y la estructura global de los contenidos y servicios disponibles en la web [BER, 2001b].

Esta semántica permitirá que agentes inteligentes puedan "entender" el significado de los documentos, con lo que se mejorará la búsqueda y recuperación de información en la web y se potenciará el desarrollo de aplicaciones en campos como el comercio electrónico, la gestión del conocimiento y la educación, la recuperación de información, entre otros. [CAS, 2003b]

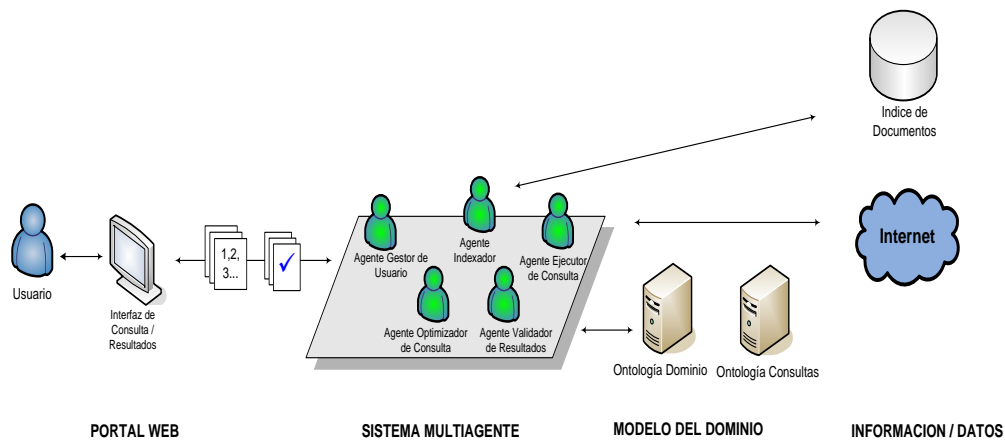
### **3.1 PROPÓSITO Y COMPONENTES DEL SISTEMA.**

El trabajo planteado, consiste de un sistema multiagente para la optimización de búsquedas que con base en el modelado del conocimiento en Mampostería Estructural, se encarga de filtrar el conjunto de páginas entregadas por un motor de búsqueda convencional, buscando un resultado mas preciso enmarcado dentro del dominio modelado. Dada una consulta, la relevancia que la información entregada al usuario tiene dentro del dominio, determina la precisión del resultado. La relevancia es el único criterio de evaluación del sistema ya que no se tienen en cuenta aspectos como tiempos de espera, tamaño del índice etc.

Los componentes del sistema son los siguientes: el Modelo del Dominio, el cual representa los aspectos más importantes de la temática seleccionada mediante una Ontología (expuesta en el capítulo 2), con la cual interactúa el sistema multiagente para determinar cuáles de las páginas devueltas por el motor de búsqueda empleado son relevantes en el dominio.

#### **3.1.1 Arquitectura del sistema**

El sistema ha sido desarrollado bajo una arquitectura de tres capas: la interacción con el usuario se maneja por medio del portal web, el cual contiene las interfaces de búsqueda y resultados; la lógica o inteligencia del sistema es manejada por medio del sistema multiagente y por último, los modelos de conocimiento y los datos son manejados por medio de ontologías y bases de datos. La figura 4., ilustra la arquitectura del sistema.



**Figura 4.** Arquitectura del sistema

El componente que brinda la “inteligencia” al sistema es el llamado Sistema Multiagente (MAS) encargado de manejar la lógica del sistema, interactúa con el usuario a través del portal web, optimiza las consultas y valida las páginas recuperadas utilizando el modelo del dominio, indizando las páginas validadas. El prototipo de portal web es el componente que contiene las interfaces de búsqueda y por medio de él se presentan los resultados a los usuarios. A través de él los clientes acceden al sistema para iniciar el sistema multiagente y realizar todos los procesos de búsqueda y recuperación de información.

### 3.2 PROCESO DE DESARROLLO DEL SISTEMA MULTIAGENTE (MAS)

Para el desarrollo del Sistema Multiagente, se analizaron algunas de las metodologías existentes para el modelado de esta clase de sistemas (ver anexos). Con base en este análisis, se determinó Gaia [WOO, 1999] como la metodología más adecuada a seguir para el proceso de modelado del MAS.

GAIA permite realizar un análisis sistemático que abarca la definición de los requerimientos del sistema. La etapa de análisis empieza desde un punto de vista abstracto: está compuesta del modelo de roles, donde se identifican los roles claves en el sistema, y el modelo de interacción, donde se definen las dependencias y relaciones entre varios roles. El proceso de diseño se basa en un modelo de agente que documenta los varios tipos de agentes, y las instancias de los mismos; un modelo de servicio que

identifica los servicios asociados con cada rol de los agentes y especifica las propiedades principales de esos servicios; y un modelo de familiaridad que define la interacción que existe entre diferentes tipos de agentes. [GAR, 2002].

El diseño producido por GAIA es generalmente abstracto, por lo que se decidió utilizar AUML<sup>4</sup>, para obtener un diseño más detallado. AUML extiende UML incluyendo un proceso de desarrollo a tres niveles, donde el primer nivel modela la funcionalidad global del sistema, un segundo nivel modelando la interacción entre agentes y un tercer nivel representando el procesamiento interno de cada agente, los cuales proporcionan mayor acercamiento a la implementación del sistema [GAR, 2002].

### 3.2.1 Especificación de requerimientos

Como primer paso en el proceso de Desarrollo del sistema, se realizó la captura de requerimientos, dando como resultado la siguiente especificación de funciones y atributos del sistema. La tabla 1., resume estos resultados.

<b>Función</b>	<b>Categoría</b>	<b>Atributo</b>	<b>Detalles y restricciones</b>	<b>Categoría</b>
Capturar la solicitud de consulta del usuario	Evidente	Metáfora de Interfaz	(detalle) Interfaz Web, maximizar fácil introducción de la consulta	Obligatoria
Presentar al usuario la ontología del dominio	Superflua	Metáfora de Interfaz	(detalle) Interfaz Web, maximizar fácil navegación entre los niveles de la ontología	Opcional
Normalizar la consulta introducida por el usuario, eliminando de ella palabras irrelevantes en el dominio.	Ocultas			
Enriquecer la consulta del usuario, adicionando, eliminando o reemplazando palabras de ella.	Ocultas			
Consultar el Índice del sistema, recuperando la información almacenada.	Ocultas			
Consultar el Motor de Búsqueda, recuperando páginas de Internet.	Ocultas			

<sup>4</sup>Extensión a UML para el diseño de sistemas basados en agentes

Generar Consultas a ser ejecutadas en el motor de búsqueda seleccionado.	Ocultas			
Revisar la validez de las páginas recuperadas, discriminando aquellas relevantes en el dominio de aquellas que no lo son.	Ocultas			
Calcular el peso de cada página para verificar su validez.	Ocultas			
Ordenar las páginas recuperadas en función de sus pesos, para presentarlas al usuario.	Ocultas			
Presentar al usuario los resultados obtenidos de la consulta	Evidente	Metáfora de Interfaz	(detalle) Interfaz Web, maximizar fácil navegación entre resultados	Obligatoria

**Tabla 1.** Resumen captura de requerimientos

### 3.2.1.1 Modelo de Casos de Uso

Inicialmente se identifican y describen los actores del sistema, se obtiene un diagrama de casos de uso el cual se ilustra en la figura 5. y se describen de manera resumida los casos de uso identificados, para adquirir una visión global de todo el sistema, dado que la metodología seleccionada para el modelado del sistema multiagente (GAIA [WOO, 1999]) es una metodología exclusiva de agentes y no es suficiente para lograr esta visión.

➤ Actores del sistema:

**Usuario:** el usuario del sistema puede ser todo tipo de personas que estén interesadas en el tema Mampostería Estructural propio del área de construcción en la Ingeniería Civil, tales como estudiantes, docentes, especialistas y trabajadores en el área. No se requiere que el usuario posea conocimiento alguno de este tópico para utilizar o acceder a los servicios de la aplicación.

**Motor de Búsqueda:** es una entidad externa al sistema, que le proporciona la información relacionada con los recursos web que tiene almacenadas.

Índice: es una entidad externa, por medio de la cual se realizan los procesos de indexación y búsqueda de los recursos web que se han recuperado y han sido validados por el sistema.

➤ Diagrama de casos de uso

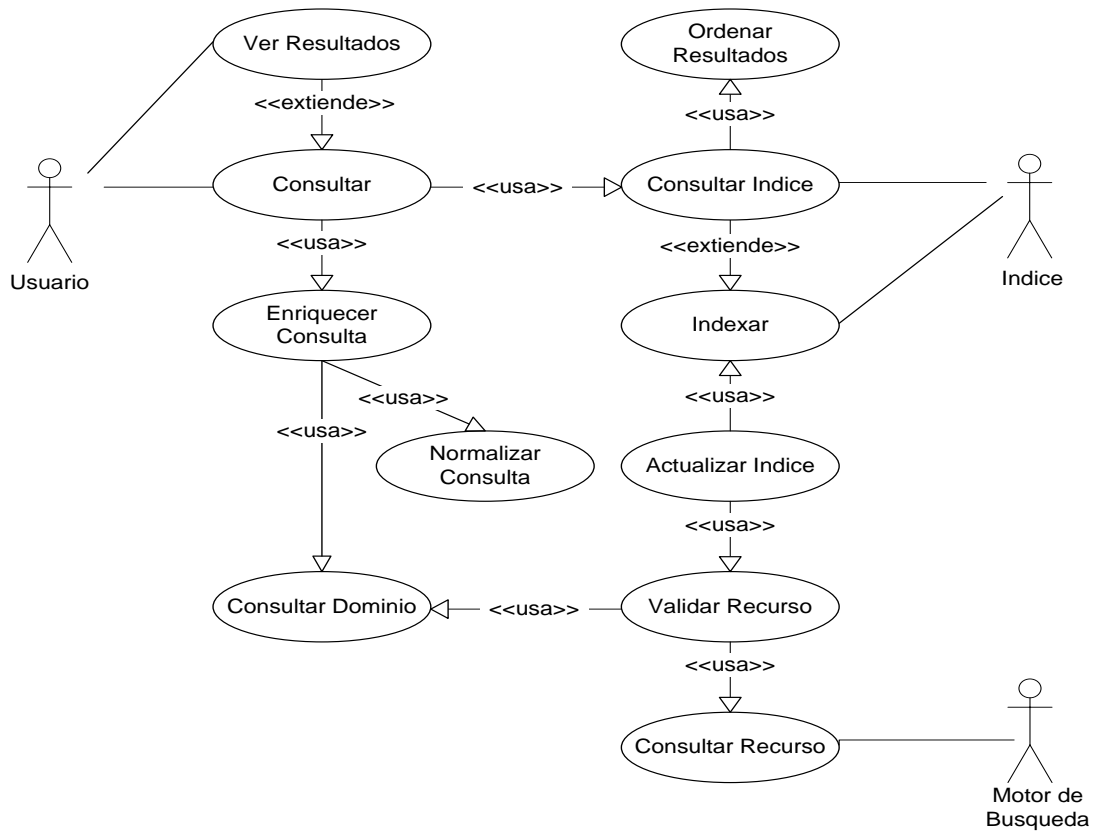


Figura 5. Diagrama de casos de uso

➤ Descripción resumida de los casos de uso

<b>Caso de uso</b>	<b>Consultar</b>
Actores	Usuario
Prioridad	Alta
Descripción	Este caso de uso se inicia cuando el usuario ingresa una consulta o elige el tema del cual desea obtener información. El propósito de este caso de uso es entregar al usuario un listado de recursos disponibles en Internet relevantes en el dominio seleccionado. Este caso de uso se relaciona con el caso de uso consultar índice y ver resultados.



<b>Caso de uso</b>	<b>Enriquecer Consulta</b>
Actores	
Prioridad	Alta
Descripción	Este caso de uso es ejecutado desde el caso de uso consultar. Utiliza los casos de uso Normalizar parámetros de búsqueda y Consultar dominio para convertir la cadena de búsqueda en un parámetro coherente a la estructura del índice.

<b>Caso de uso</b>	<b>Normalizar Consulta</b>
Actores	
Prioridad	Alta
Descripción	Este caso de uso se ejecuta desde el caso Enriquecer Consulta. Modifica la consulta inicial del usuario, eliminando de ella los términos irrelevantes en el dominio.

<b>Caso de uso</b>	<b>Consultar dominio</b>
Actores	
Prioridad	Alta
Descripción	Este caso de uso se ejecuta desde los casos Enriquecer Consulta y Validar Recurso. Proporciona información concerniente al modelo del dominio.

<b>Caso de uso</b>	<b>Consultar índice</b>
Actores	Índice
Prioridad	Alta
Descripción	Este caso de uso se ejecuta desde el caso de uso Consultar. Se ejecutan las consultas de usuario sobre el índice y se retornan las referencias a los recursos relacionados con la temática buscada.

<b>Caso de uso</b>	<b>Consultar Recurso</b>
Actores	Motor de Búsqueda
Prioridad	Alta
Descripción	Este caso de uso se ejecuta desde el caso de uso Validar Recurso. Proporciona mecanismos para enviar al motor de búsqueda, diferentes consultas con el fin de recuperar recursos relacionados con la temática modelada.

<b>Caso de uso</b>	<b>Actualizar Índice</b>
Actores	
Prioridad	Alta
Descripción	Cada cierto periodo de tiempo, el sistema inicia este procedimiento para actualizar el índice y para consultar nuevos recursos en Internet; para ello utiliza los casos de uso Consultar Recurso y Validar Recurso.

<b>Caso de uso</b>	<b>Indexar</b>
Actores	Índice
Prioridad	Alta
Descripción	Este caso de uso se ejecuta desde Actualizar índice. Su propósito es adicionar al índice cada uno de los recursos consultados y validados.

<b>Caso de uso</b>	<b>Validar Recurso</b>
Actores	
Prioridad	Alta
Descripción	Este caso de uso se ejecuta desde el caso de uso Actualizar Índice. Su propósito es establecer la relevancia del recurso con el modelo del dominio.

<b>Caso de uso</b>	<b>Obtener Recurso</b>
Actores	
Prioridad	Alta
Descripción	Este caso de uso se ejecuta desde el caso de uso Actualizar índice. El propósito de este caso de uso es realizar consultas en Internet sobre recursos relacionados con el modelo del dominio. Adecua el tipo de recurso encontrado en un formato apto para el caso de uso Validar Recurso.

<b>Caso de uso</b>	<b>Ver resultados</b>
Actores	Usuario
Prioridad	Alta
Descripción	Este caso de uso se extiende del caso de uso buscar. Una vez ejecutado el caso de uso Consultar Índice, se presenta al usuario una interfaz con el resultado de la búsqueda.

### 3.2.2 Análisis

- Modelo de roles

La tabla 2., describe el modelo de roles de los agentes del sistema

Departamento	Roles	Descripción
<b>Gestión de Usuario</b>	Gestor de Presentación	Tiene como función crear y mostrar al usuario las interfaces inicial y de resultados de la consulta.
<b>Recuperación de Información</b>	Normalizador de Consulta inicial	Tiene como función Normalizar la consulta inicial formulada por el usuario, antes de ser enriquecida.
	Enriquecedor de Consulta Normalizada	Es el encargado de enriquecer la consulta utilizando la información del dominio
	Ejecutor de Consulta en Motor de Búsqueda.	Es el encargado de ejecutar la consulta enriquecida sobre el motor de búsqueda.
	Ejecutor de Consulta en Índice	Es el encargado de ejecutar la consulta enriquecida sobre el índice del sistema.
<b>Organización de Información</b>	Validador de Resultados	Tiene como función validar cuales de las páginas recuperadas, son relevantes en el dominio.
	Ordenador de Resultados	Ordenar el resultado de la consulta, garantizando que se ubiquen en primer lugar, aquellos que sean más relevantes según el dominio

<b>Indexación</b>	Gestor de Índice	Encargado de manejar el índice del sistema, indexando las páginas validadas previamente.
	Gestor de Actualización	Encargado de actualizar periódicamente el índice, registrando las consultas no disponibles en el sistema y buscando periódicamente nuevos recursos para satisfacer dichas consultas
<b>Localización</b>	Gestor de Directorio	Tiene como función mantener la información de los agentes, localizarlos y facilitar la comunicación entre ellos.

**Tabla 2.** Modelo de Roles

➤ Modelo de Rol detallado

<b>Esquema del Rol: <i>Gestor de Presentación (GP)</i></b>	
<b>Descripción:</b> Tiene como función crear en forma personalizada al usuario las interfaces inicial y de resultados de la consulta.	
<b>Protocolos y Actividades:</b> peticionInterfazConsulta peticionInterfazResultados crearInterfazConsulta crearInterfazResultados	
<b>Permisos:</b>	
• <b>Leer:</b>	
<b>informado</b> perfilUsuario	
<b>informado</b> resultadoOrdenamiento	// Páginas validadas, ordenadas según su relevancia en el dominio
• <b>Generar:</b>	
interfazConsulta	
interfazPresentacionResultados	//Interfaces necesarias para la interacción del sistema con el usuario.
• <b>Modificar:</b>	
<b>Responsabilidades</b>	
• <b>Ejecución:</b>	GP = (INICIAR.RESULTADO) INICIAR = PeticionInterfazConsulta.CrearInterfazConsulta RESULTADO = PeticionInterfazResultados.CrearInterfazResultados)
• <b>Seguridad:</b>	idUsuario ≠ null

<b>Esquema del Rol: <i>Asistente de Usuario (AU)</i></b>	
<b>Descripción:</b> Es el encargado de hacer el seguimiento al usuario, asistiéndole en el proceso de realizar la consulta	
<b>Protocolos y Actividades:</b> IdentificarUsuario CapturarConsulta NavegarOntologia RegistrarInteresUsuario GestionarHistorial EnviarPeticonConsulta	
<b>Permisos:</b>	
• <b>Leer:</b>	idUsuario // Identificación del Usuario
• <b>informado</b>	consultalncial // Consulta inicial introducida por el usuario
• <b>Generar:</b>	temasRelacionados //Posibles consultas de acuerdo con los niveles de la ontología
• <b>Modificar:</b>	interesesUsuarios historialConsulta
<b>Responsabilidades</b>	
• <b>Ejecución:</b>	AU = IDENTIFICAR.(ASISTIR.CONSULTAR.(APRENDER)*) $\omega$ IDENTIFICAR = IdentificarUsuario ASISTIR = (NavegarOntologia)*   CapturarConsulta CONSULTAR = EnviarPeticonConsulta APRENDER = GestionarHistorial.RegistrarInteresUsuario
<b>Seguridad</b>	consultalncial $\neq$ null

<b>Esquema del Rol: <i>Normalizador de Consulta Inicial (NCI)</i></b>	
<b>Descripción:</b> Tiene como función Normalizar la consulta inicial formulada por el usuario, antes de ser enriquecida.	
<b>Protocolos y Actividades:</b> PeticonNormalizacion ExtraerPalabras EliminarIrrelevantes ResultadoNormalizacion	
<b>Permisos:</b>	
• <b>Leer:</b>	idUsuario // Identificación del Usuario
• <b>informado</b>	consultalncial // Consulta inicial introducida por el usuario
• <b>informado</b>	palabrasIrrelevantes // Repositorio de palabras que no deben utilizarse en la consulta
• <b>Generar:</b>	consultaNormalizada // Cadena de Búsqueda normalizada. Palabras clave en la consulta del usuario
• <b>Modificar:</b>	
<b>Responsabilidades:</b>	
<b>Ejecución:</b>	NCI = (RECIBIR.NORMALIZAR.ENTREGAR) $\omega$ RECIBIR = PeticonNormalizacon NORMALIZAR=(ExtraerPalabra.CompararListado.EliminarIrrelevantes) + ENTREGAR = ResultadoNormalizacion
<b>Seguridad:</b>	consultalncial $\neq$ null

<b>Esquema del Rol: <i>Enriquecedor de Consulta Normalizada (ECN)</i></b>										
<b>Descripción:</b> Es el encargado de enriquecer la consulta utilizando la información del dominio										
<b>Protocolos y Actividades:</b> SolicitudEnriquecerConsulta EliminarPalabras, AdicionarPalabras, ReemplazarPalabras ResultadoEnriquecerConsulta										
<b>Permisos:</b>										
• <b>Leer:</b>	<table> <tr> <td>informado</td> <td>modeloDominio</td> <td>// información del Dominio</td> </tr> <tr> <td>informado</td> <td>idUsuario</td> <td>// Identificación del Usuario</td> </tr> <tr> <td>informado</td> <td>consultaNormalizada</td> <td>// Cadena de Búsqueda normalizada. Palabras clave en la consulta del usuario</td> </tr> </table>	informado	modeloDominio	// información del Dominio	informado	idUsuario	// Identificación del Usuario	informado	consultaNormalizada	// Cadena de Búsqueda normalizada. Palabras clave en la consulta del usuario
informado	modeloDominio	// información del Dominio								
informado	idUsuario	// Identificación del Usuario								
informado	consultaNormalizada	// Cadena de Búsqueda normalizada. Palabras clave en la consulta del usuario								
• <b>Generar:</b>	<table> <tr> <td>consultaEnriquecida</td> <td></td> <td>// Cadena de Búsqueda optimizada. Palabras clave para ejecutar la consulta</td> </tr> </table>	consultaEnriquecida		// Cadena de Búsqueda optimizada. Palabras clave para ejecutar la consulta						
consultaEnriquecida		// Cadena de Búsqueda optimizada. Palabras clave para ejecutar la consulta								
• <b>Modificar:</b>										
<b>Responsabilidades:</b>										
<b>Ejecución:</b>	<p>ECN = (RECIBIR.ENRIQUECER.ENTREGAR)<math>\omega</math>  RECIBIR = SolicitudEnriquecerConsulta  NORMALIZAR = (ExtraerPalabra+.EliminarPalabra*.  AdicionarPalabra*.ReemplazarPalabra* ) +  ENTREGAR = ResultadoEnriquecerConsulta</p>									
<b>Seguridad</b>	consultaNormalizada $\neq$ null									

<b>Esquema del Rol: <i>Ejecutor de Consulta en Motor (ECM)</i></b>							
<b>Descripción:</b> Es el encargado de ejecutar la consulta enriquecida sobre el motor de búsqueda para recuperar los recursos disponibles en Internet para ser validados.							
<b>Protocolos y Actividades:</b> SolicitudEjecucionConsulta ConsultarMotor EntregarResultadoEjecucion							
<b>Permisos:</b>							
• <b>Leer:</b>	<table> <tr> <td>informado</td> <td>idUsuario</td> <td>// Identificación del Usuario</td> </tr> <tr> <td>informado</td> <td>consultaEnriquecida</td> <td>// Cadena de Búsqueda optimizada. Palabras clave para ejecutar la consulta</td> </tr> </table>	informado	idUsuario	// Identificación del Usuario	informado	consultaEnriquecida	// Cadena de Búsqueda optimizada. Palabras clave para ejecutar la consulta
informado	idUsuario	// Identificación del Usuario					
informado	consultaEnriquecida	// Cadena de Búsqueda optimizada. Palabras clave para ejecutar la consulta					
• <b>Generar:</b>	<table> <tr> <td>resultadoConsulta</td> <td></td> <td>// Páginas que contienen las palabras clave de la consulta</td> </tr> </table>	resultadoConsulta		// Páginas que contienen las palabras clave de la consulta			
resultadoConsulta		// Páginas que contienen las palabras clave de la consulta					
• <b>Modificar:</b>							
<b>Responsabilidades:</b>							
<b>Ejecución:</b>	<p>ECE = (RECIBIR.CONSULTAR.ENTREGAR)<math>\omega</math>  RECIBIR = RecibirConsultaEnriquecida  CONSULTAR = ConsultarMotor  ENTREGAR = EnviarResultadoConsulta</p>						
<b>Seguridad</b>	ConsultaEnriquecida $\neq$ null						

<b>Esquema del Rol: <i>Ejecutor de Consulta en Índice (ECI)</i></b>	
<b>Descripción:</b> Es el encargado de ejecutar la consulta enriquecida sobre el índice del sistema para presentar resultados al usuario	
<b>Protocolos y Actividades:</b> SolicitudEjecucionConsulta ConsultarIndice EntregarResultadoEjecucion	
<b>Permisos:</b>	
• <b>Leer:</b>	informado idUsuario // Identificación del Usuario informado consultaEnriquecida // Cadena de Búsqueda optimizada. Palabras clave para ejecutar la consulta
• <b>Generar:</b>	resultadoConsulta // Páginas que contienen las palabras clave de la consulta
• <b>Modificar:</b>	
<b>Responsabilidades:</b>	
<b>Ejecución:</b>	ECE = (RECIBIR.CONSULTAR.ENTREGAR)ω RECIBIR = RecibirConsultaEnriquecida CONSULTAR = ConsultarIndice ENTREGAR = EnviarResultadoConsulta
<b>Seguridad</b>	ConsultaEnriquecida ≠ null

<b>Esquema del Rol: <i>Validador de Resultados (VR)</i></b>	
<b>Descripción:</b> Tiene como función validar cuales de las páginas recuperadas, son relevantes en el dominio.	
<b>Protocolos y Actividades:</b> SolicitudValidacion ExtraerPagina ConsultarDominio CalcularPeso RevisarValidez EntregarResultadoValidación	
<b>Permisos:</b>	
• <b>Leer:</b>	informado idUsuario // Identificación del Usuario informado resultadoConsulta // Páginas devueltas al ejecutar la consulta. modeloDominio // información del Dominio
• <b>Generar:</b>	resultadoValidacion // Páginas relevantes en el dominio
• <b>Modificar:</b>	
<b>Responsabilidades:</b>	
<b>Ejecución:</b>	VR = (RECIBIR.VALIDAR.ENTREGAR)ω RECIBIR = SolicitudValidacion VALIDAR = (ExtraerPagina.ConsultarDominio.CalcularPeso. RevisarValidez ) + ENTREGAR = EntregarResultadoValidacion
<b>Seguridad</b>	resultadoConsulta ≠ null

<b>Esquema del Rol: Ordenador de Resultados (OR)</b>	
<b>Descripción:</b> Ordenar el resultado de la consulta, garantizando que se ubiquen en primer lugar, aquellos que sean más relevantes según el dominio	
<b>Protocolos y Actividades:</b> SolicitarOrdenamiento ExtraerPagina ObtenerPeso OrdenarPesos EntregarResultadoOrdenamiento	
<b>Permisos:</b>	
<ul style="list-style-type: none"> <li>• <b>Leer:</b> <ul style="list-style-type: none"> <li>informado idUsuario // Identificación del Usuario</li> <li>informado resultadoValidacion // Páginas devueltas al ejecutar la validacion.</li> <li>informado pesoPagina // Peso de cada página obtenido en la validacion</li> </ul> </li> <li>• <b>Generar:</b> resultadoOrdenamiento // Páginas validadas ordenadas según su relevancia en el dominio</li> <li>• <b>Modificar:</b></li> </ul>	
<b>Responsabilidades:</b>	
<b>Ejecución:</b>	OR = (RECIBIR.ORDENAR.ENTREGAR) $\omega$ RECIBIR = SolicitudOrdenamiento VALIDAR = (ExtraerPagina.ObtenerPeso) +.OrdenarPesos ENTREGAR = EntregarResultadoOrdenamiento
<b>Seguridad</b>	resultadoValidacion $\neq$ null pesoPagina $\neq$ null

<b>Esquema del Rol: Gestor de Índice (GI)</b>	
<b>Descripción:</b> Encargado de crear el índice de recursos válidos en el dominio existentes en Internet	
<b>Protocolos y Actividades:</b> ConsultarDominio GenerarConsultas SolicitudEjecucionConsulta SolicitudValidacion AdicionarPaginasValidadas	
<b>Permisos:</b>	
<ul style="list-style-type: none"> <li>• <b>Leer:</b> <ul style="list-style-type: none"> <li>informado resultadoValidacion // Páginas relevantes en el dominio</li> <li>informado modeloDominio // información del Dominio</li> </ul> </li> <li>• <b>Generar:</b> indice // Índice inicial basado en el modelo de dominio</li> <li>• <b>Modificar:</b> indiceTemp // Páginas validadas previamente para adicionar al índice en la siguiente actualización índice // Modificado en cada actualización</li> </ul>	
<b>Responsabilidades:</b>	
<b>Ejecución:</b>	GI = CREAR.(ACTUALIZAR) $\omega$ CREAR = ConsultarDominio.BUSCARPAGINAS ACTUALIZAR=ConsultarHistorial.BUSCARPAGINAS BUSCARPAGINAS = (GenerarConsultas.SolicitudEjecucionConsulta. SolicitudValidacion. AdicionarPaginasValidadas) +
<b>Seguridad</b>	true

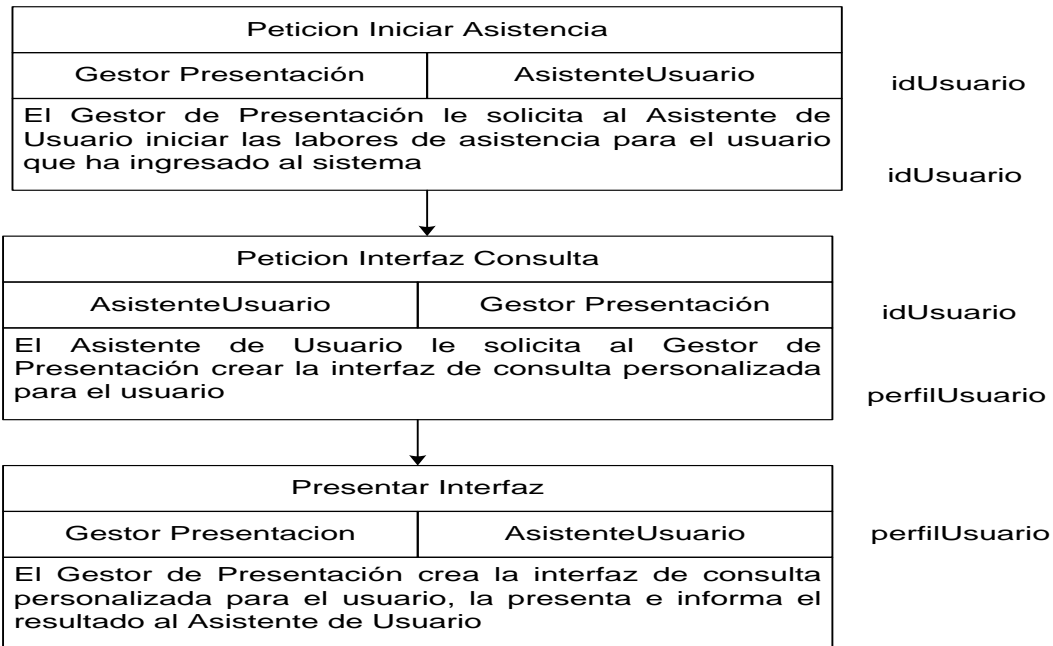
<b>Esquema del Rol:</b> <i>Gestor de Actualización (GA)</i>
<b>Descripción:</b> Tiene como función mantener actualizado el índice, registrando consultas válidas en el dominio que aún no tienen solución en el mismo y activando periódicamente el proceso de actualización del Índice de recursos en el sistema.
<b>Protocolos y Actividades:</b> SolicitudActualizacion <u>AdicionarNuevaConsulta</u> SolicitudNuevaConsulta InformarNuevaConsulta
<b>Permisos:</b> <ul style="list-style-type: none"> <li>• <b>Leer:</b> informado nuevaConsulta // consulta válida en el dominio que aún no tiene solución en el sistema tiempoActualizacion // información del tiempo transcurrido desde la última actualización del índice</li> <li>• <b>Generar:</b></li> <li>• <b>Modificar:</b></li> </ul>
<b>Responsabilidades:</b>  <b>Ejecución:</b> GA = (AdicionarNuevaConsulta)ω (SolicitudActualizacion.SolicitudNuevaConsulta.InformarNuevaConsulta)ω <b>Seguridad</b> nuevaConsulta ≠ null

<b>Esquema del Rol:</b> <i>Gestor de Directorio (GD)</i>
<b>Descripción:</b> Tiene como función mantener la información de los agentes, localizarlos y facilitar la comunicación entre ellos.
<b>Protocolos y Actividades:</b> SolicitudLocalizacion LocalizarAgente ResultadoLocalizacion
<b>Permisos:</b> <ul style="list-style-type: none"> <li>• <b>Leer:</b> informado tipoAgente // Tipo de Agente requerido directorioAgentes // información de los Agentes disponibles</li> <li>• <b>Generar:</b></li> <li>• <b>Modificar:</b></li> </ul>
<b>Responsabilidades:</b>  <b>Ejecución:</b> GD = (SolicitudLocalizacion.LocalizarAgente.ResultadoLocalizacion)ω <b>Seguridad</b> tipoAgente ≠ null directorioAgentes ≠ vacio

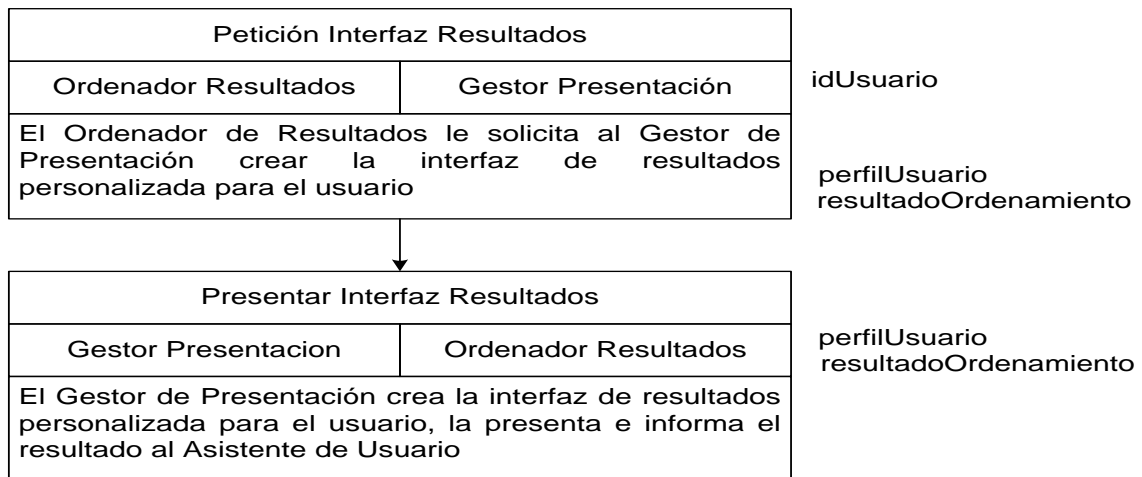


➤ Modelo de interacción

Las figuras 6. y 7., ilustran la definición de los protocolos asociados al rol Gestor de Presentación.

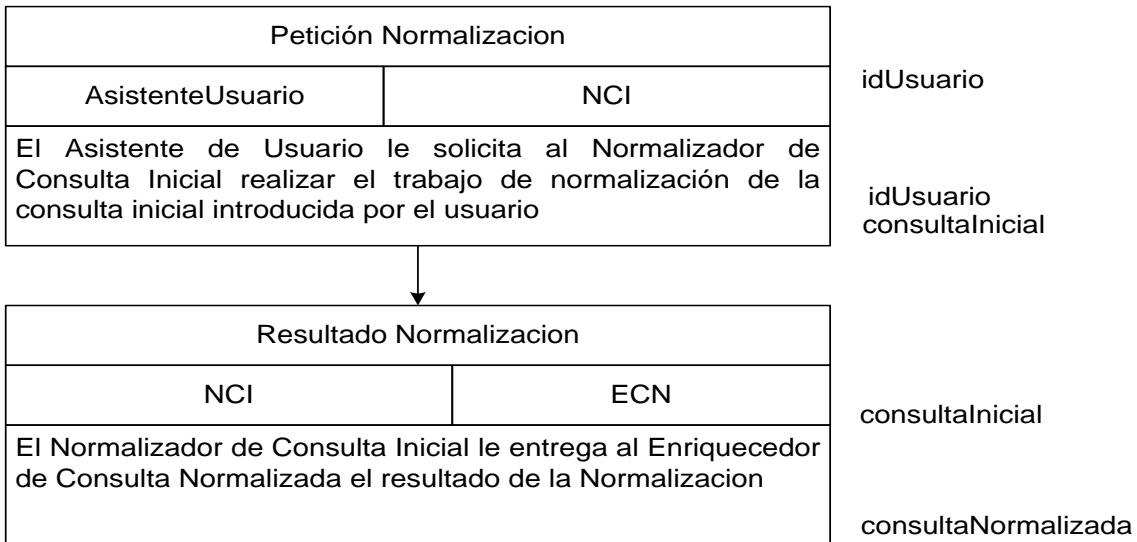


**Figura 6.** Presentar Interfaz de Consulta



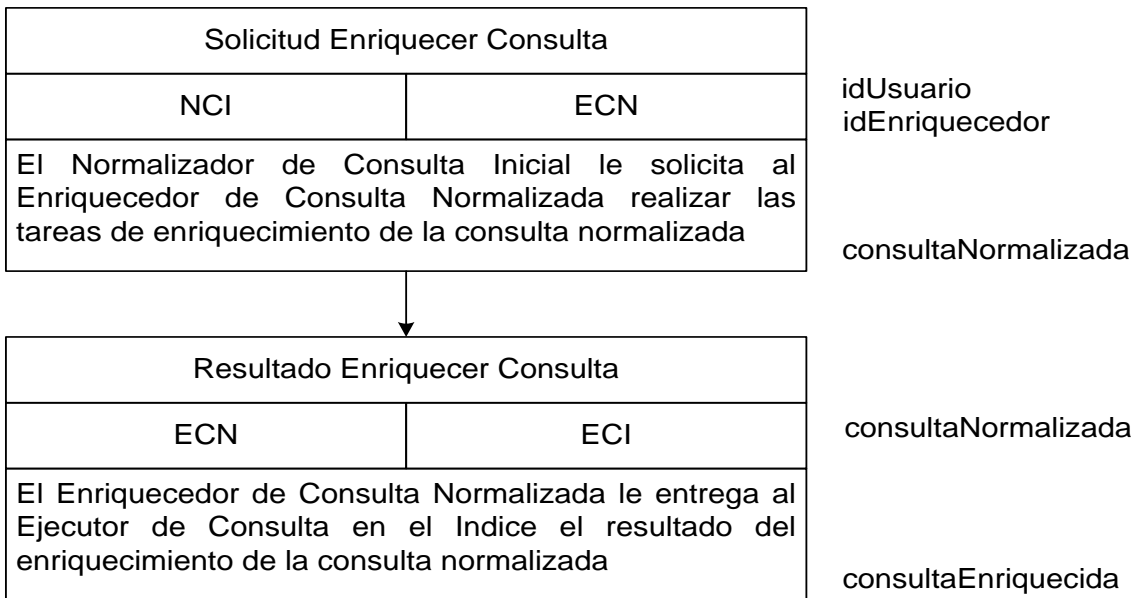
**Figura 7.** Presentar Interfaz de Resultados

La figura 8., ilustra definición de protocolos asociados al Rol Normalizador de Consulta Inicial



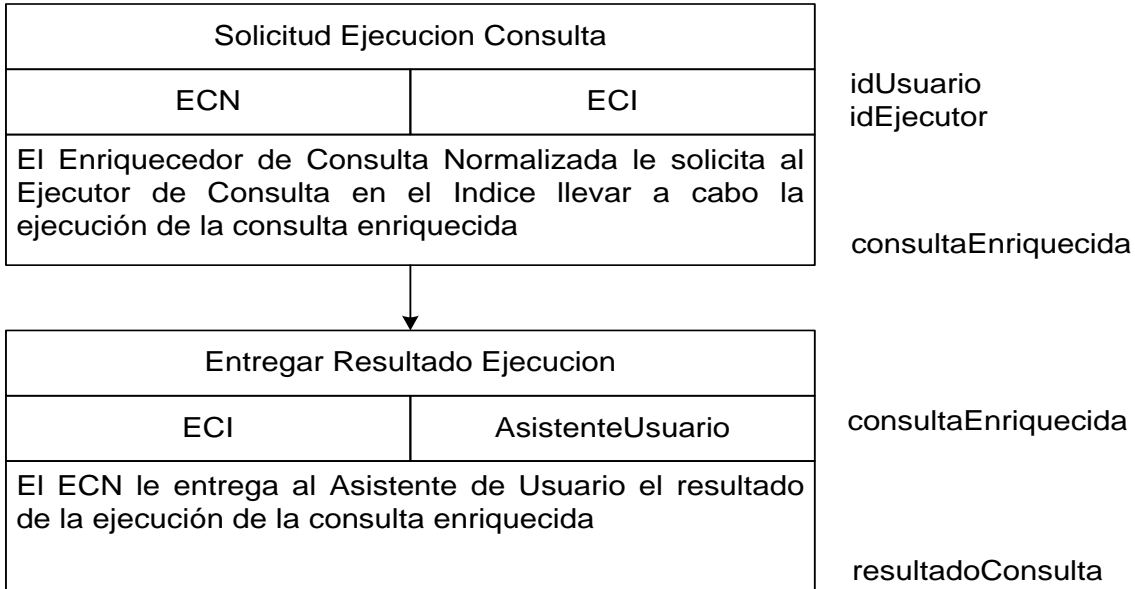
**Figura 8.** Protocolo asociado al Rol Normalizador de Consulta Inicial

La figura 9., ilustra la definición de protocolos asociados al rol Enriquecedor de Consulta Normalizada



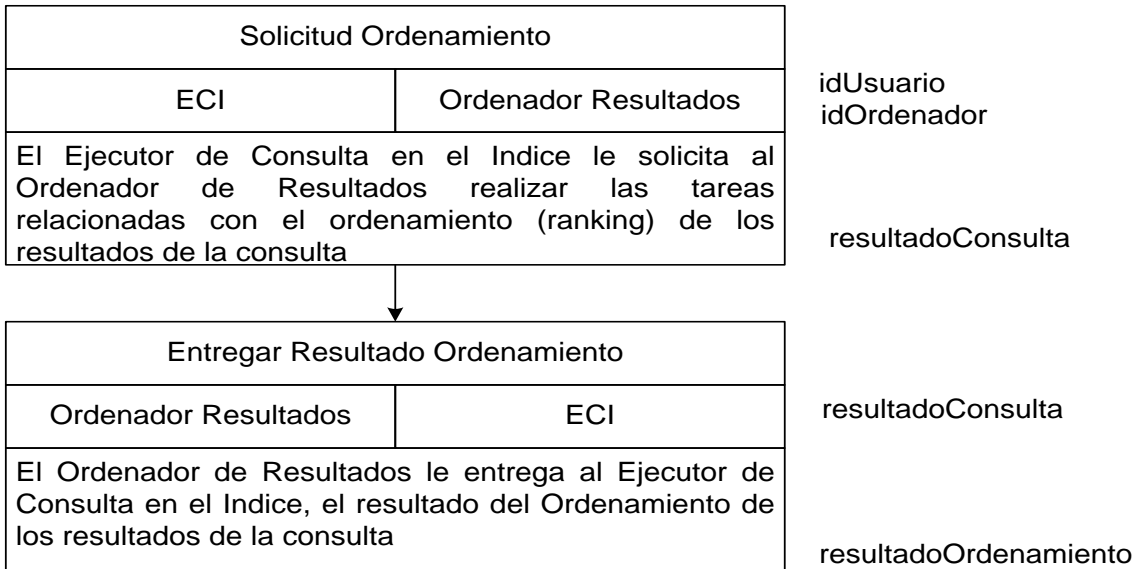
**Figura 9.** Definición protocolos asociados al rol Enriquecedor de Consulta Normalizada

La figura 10., ilustra la definición de Protocolos asociados al Rol Ejecutor de Consulta Enriquecida.



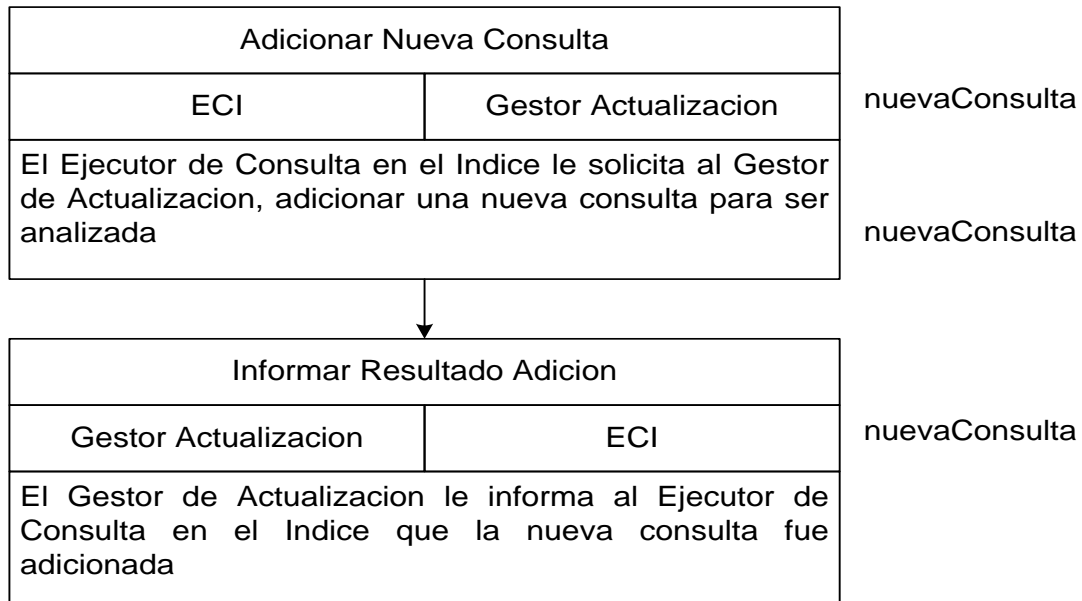
**Figura 10.** Protocolos asociados al Rol Ejecutor de Consulta Enriquecida.

La figura 11., ilustra la definición de protocolos asociados al rol Ordenador de Resultados



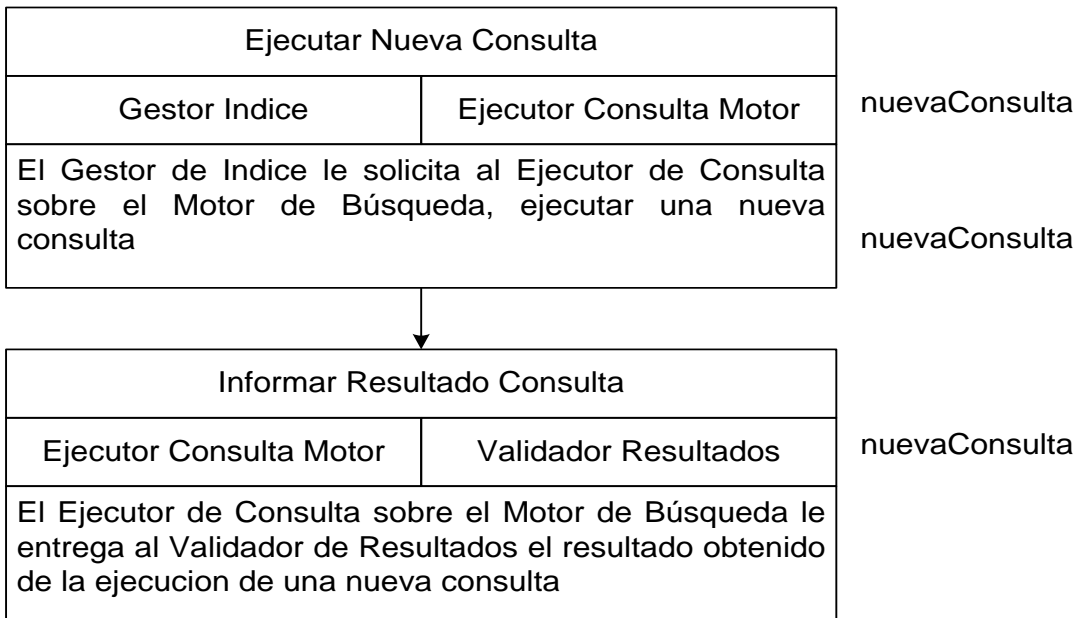
**Figura 11.** Definición de protocolos asociados al rol Ordenador de Resultados

La figura 12., ilustra la definición del protocolo asociado al rol Gestor de Actualización



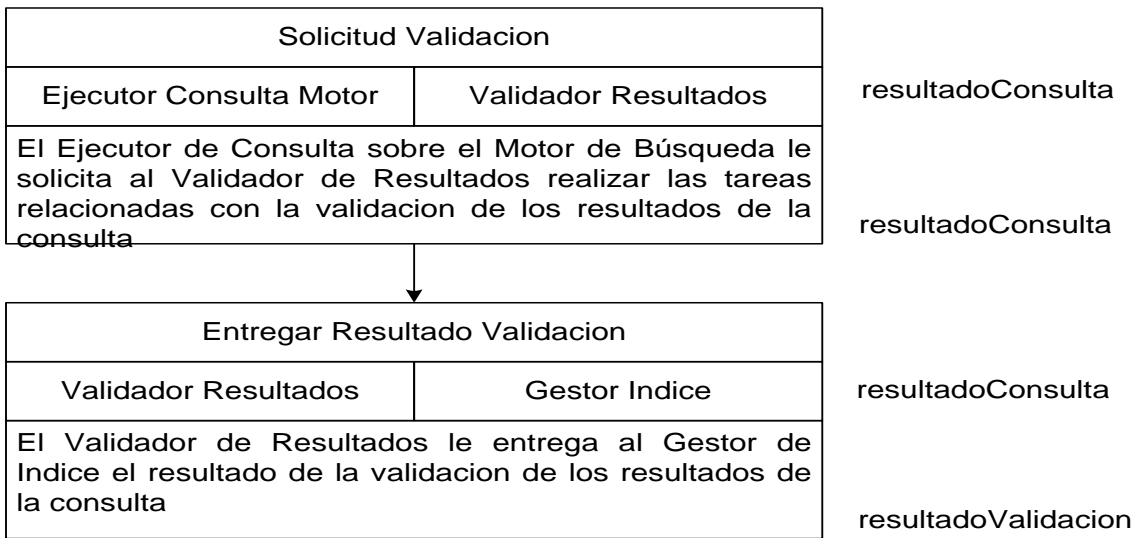
**Figura 12.** Protocolos asociados al rol Gestor de Actualización

La figura 13., ilustra la definición de protocolos asociados al rol Gestor de Índice.



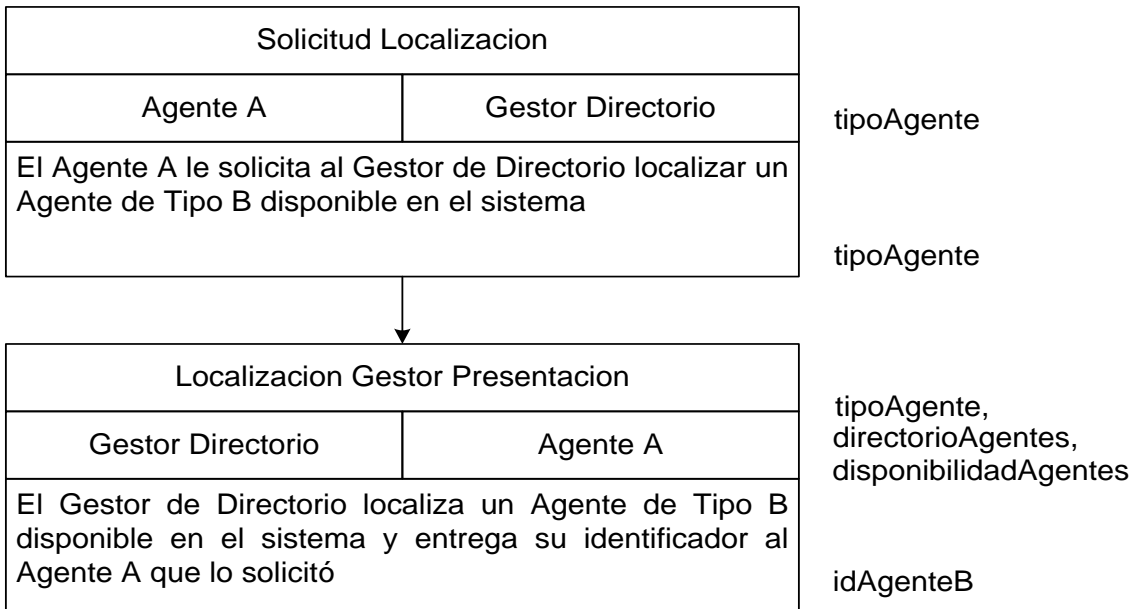
**Figura 13.** Protocolo asociado al rol Gestor de Índice

La figura 14., ilustra la definición del protocolo asociado al rol Validador de Resultados



**Figura 14.** Protocolo asociado al rol Validador de Resultados

.La figura 15., ilustra la definición del protocolo asociado al rol Gestor de Directorio



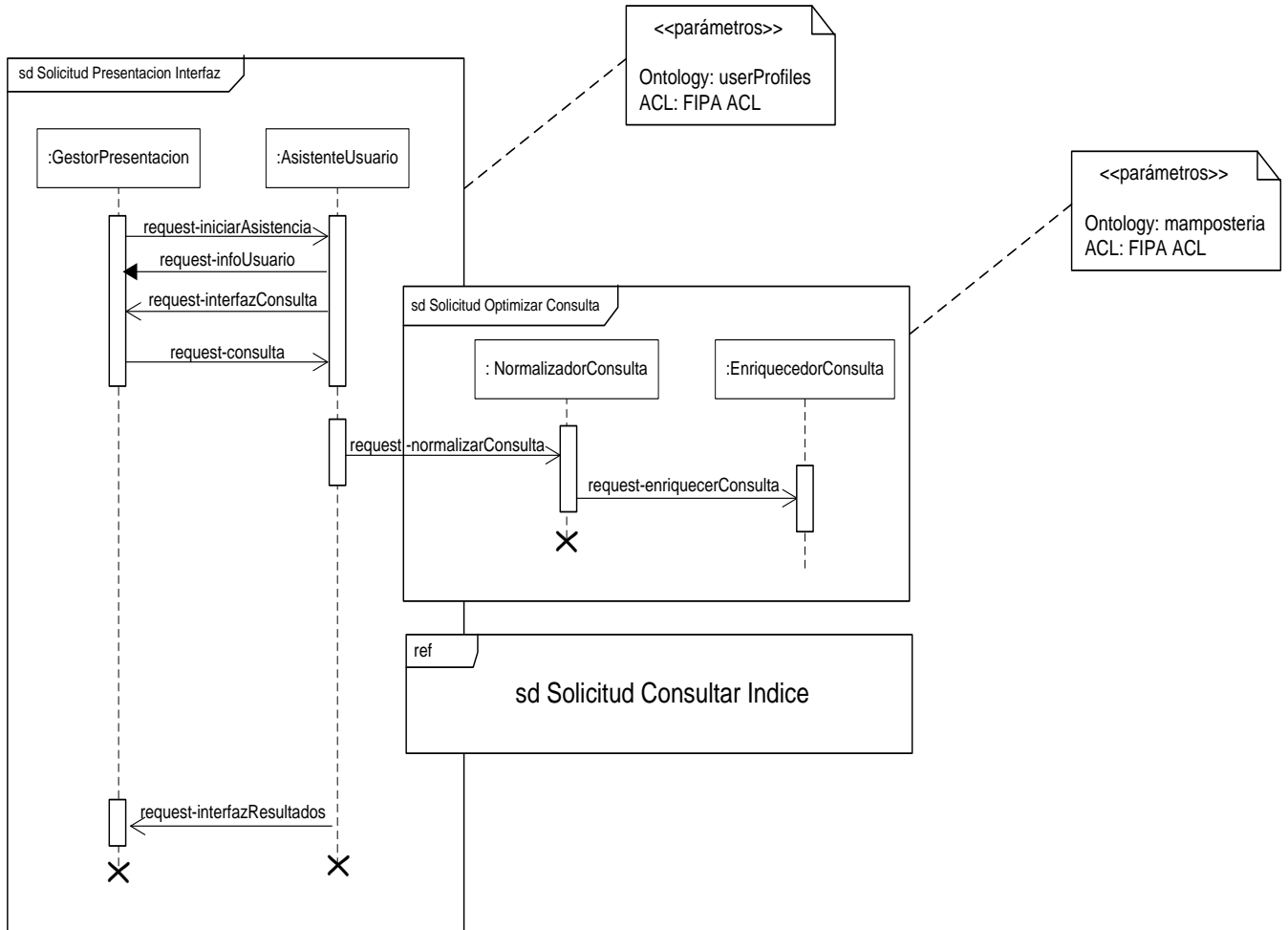
**Figura 15.** Protocolo asociado al rol Gestor de Directorio

Este protocolo se lleva a cabo cada vez que un Agente del Sistema necesita interactuar con un Agente de otro tipo:

El Agente A le solicita al Gestor de Directorio Localizar un Agente de tipo B que se encuentre disponible en el Sistema, el Gestor de Directorio lo localiza y devuelve al Agente A el identificador del Agente B. La comunicación posterior entre los dos agentes se efectúa en forma directa.

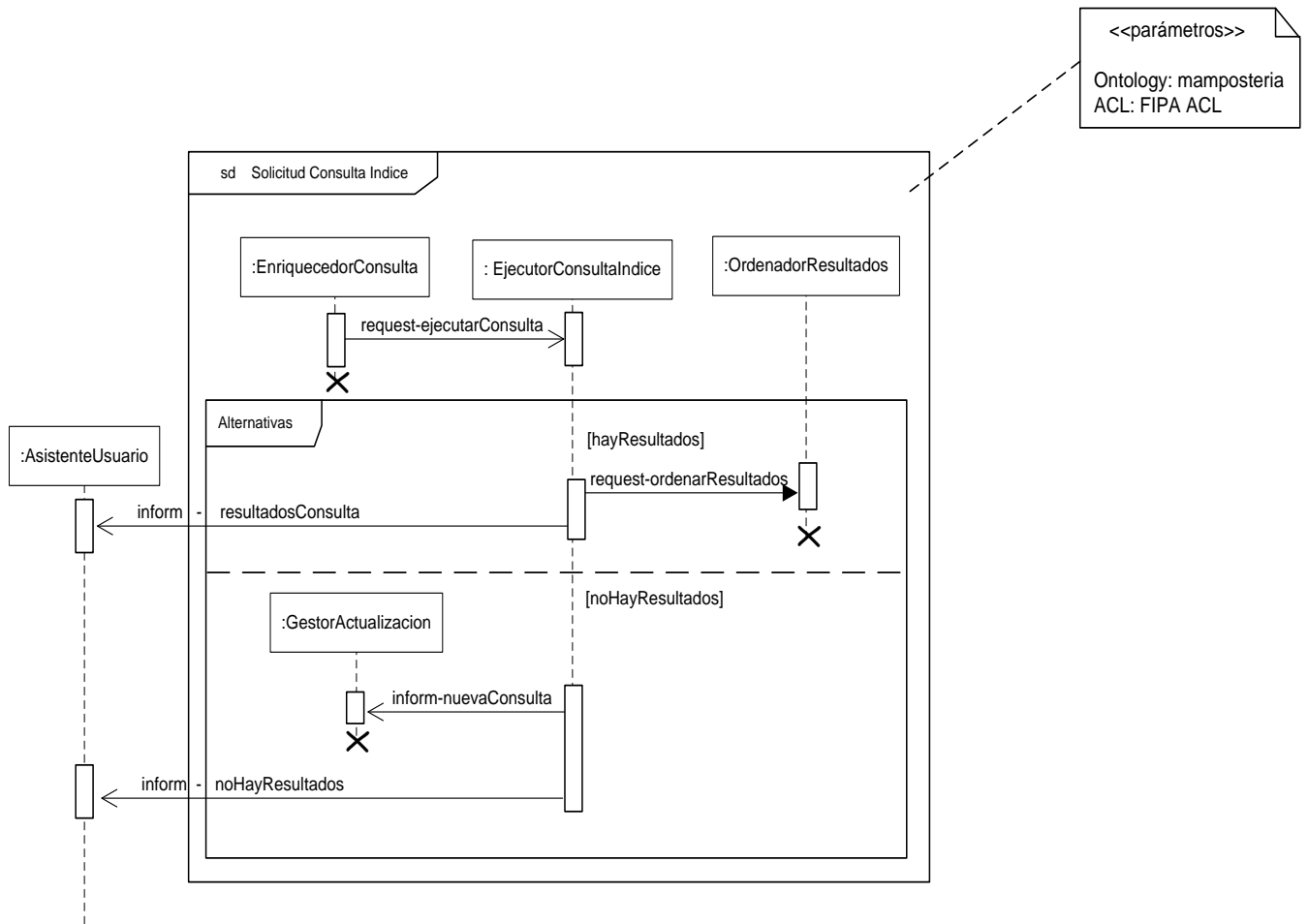
➤ Diagramas de Secuencia

La figura 16., ilustra el diagrama de secuencia Consultar Recursos en el Sistema



**Figura 16.** Diagrama de Secuencia Consultar Recursos en el Sistema

La figura 17., ilustra el diagrama de secuencia Consultar Índice



**Figura 17.** Diagrama de Secuencia Consultar Índice

La figura 18., ilustra el diagrama de secuencia Actualizar Índice

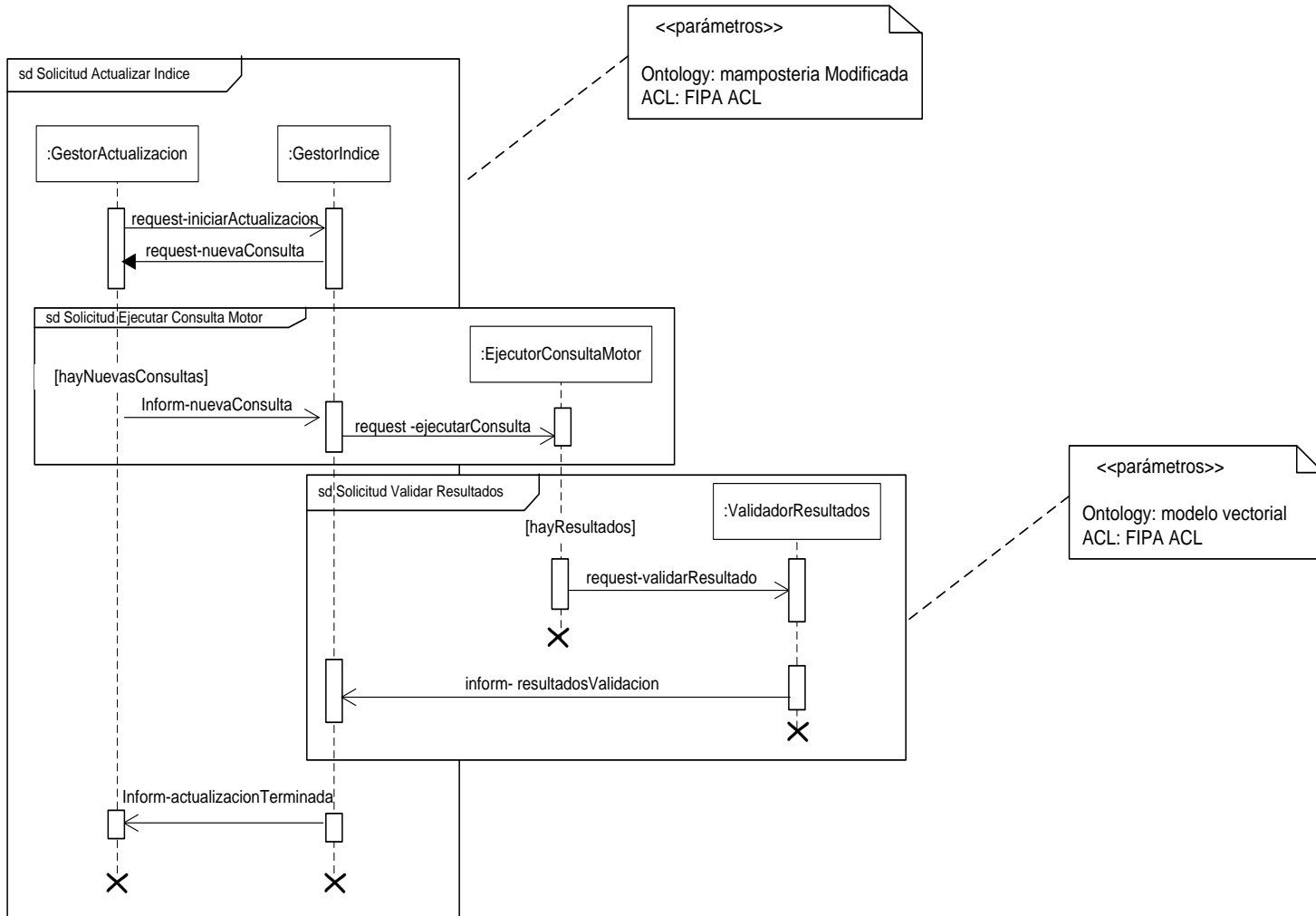


Figura 18. Diagrama de Secuencia Consultar Índice

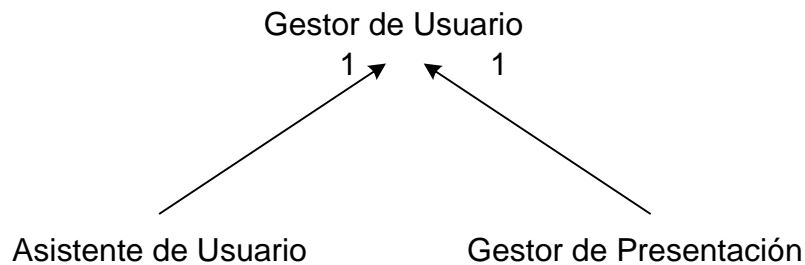


### 3.2.3 Diseño

El proceso de diseño se basa en un modelo de agente que documenta los varios tipos de agentes y las instancias de los mismos; un modelo de servicio que identifica los servicios asociados con cada rol de los agentes y especifica las propiedades principales de esos servicios y un modelo de familiaridad o conocimiento que define la interacción que existe entre diferentes tipos de agentes [GAR, 2002].

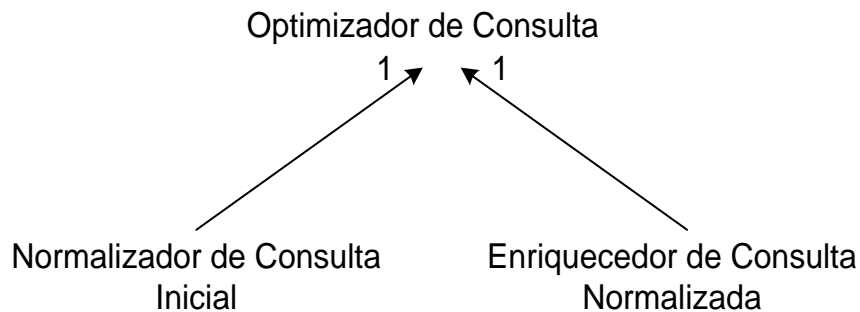
➤ **Modelo de Agentes:**

La figura 19., ilustra el modelo de agentes para el agente gestor de usuario.



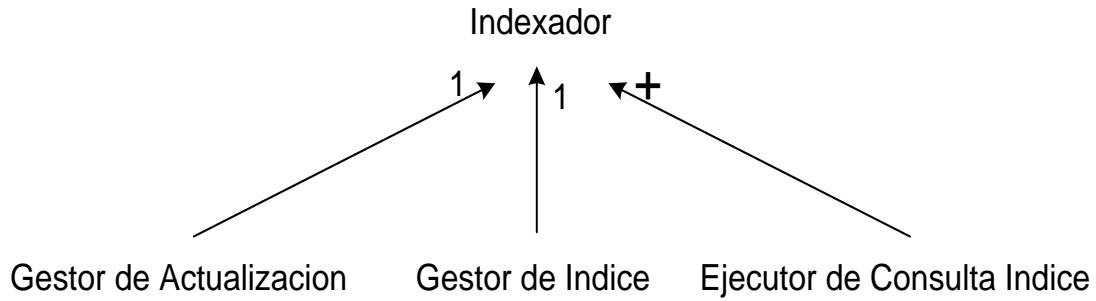
**Figura 19.** Modelo de Agentes para el Agente Gestor de Usuario.

La figura 20., ilustra el modelo de agentes para el agente optimizador de consulta.



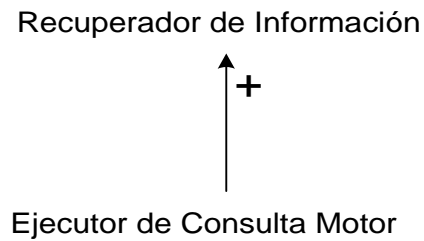
**Figura 20.** Modelo de Agentes para el Agente Optimizador de Consulta.

La figura 21., ilustra el modelo de agentes para el agente Gestor de índice.



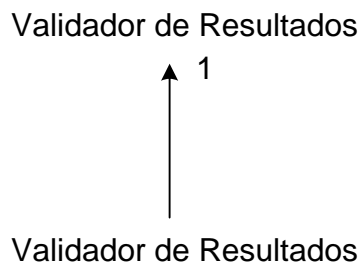
**Figura 21.** Modelo de Agentes para el Agente Gestor de índice.

La figura 22., ilustra el modelo de agentes para el agente ejecutor de consulta



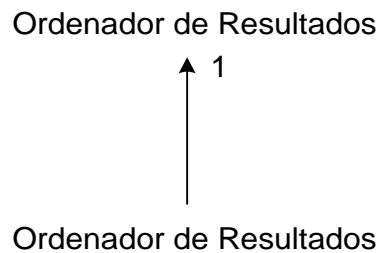
**Figura 22.** Modelo de Agentes para el agente ejecutor de consulta

La figura 23., ilustra el modelo de agentes para el agente de Validación.



**Figura 23.** Modelo de Agentes para el Agente de Validación.

La figura 24., ilustra el modelo de agentes para el agente Ejecutor de Consulta.



**Figura 24.** Modelo de Agentes para el agente Ejecutor de Consulta.

La figura 25., ilustra el modelo de agentes para el Agente Localizador.



**Figura 25.** Modelo de Agentes para el agente Localizador.

➤ Modelo de Servicios:

En la tabla 3., se describe el Modelo de Servicios asociados al rol Gestor de Presentación

Servicio	Entradas	Salidas	Pre-condiciones	Post-condiciones
Petición Interfaz Consulta		Interfaz de Consulta		true
Petición Interfaz Resultados	ResultadoOrdenamiento	Interfaz de Resultados	ResultadoOrdenamiento ≠ null	true

**Tabla 3.** Modelo de Servicios asociado al rol Gestor de Presentación

En la tabla 4., se describe el Modelo de Servicios asociados al rol Normalizador de Consulta Inicial

Servicio	Entradas	Salidas	Pre-condiciones	Post-condiciones
Petición Normalización				
Resultado Normalización	ConsultaInicial	ConsultaNormalizada	idUsuario ≠ null consultaInicial ≠ null	consultaNormalizada ≠ null

**Tabla 4.** Modelo de Servicios asociado al rol Normalizador de Consulta Inicial.

En la tabla 5., se describe el Modelo de Servicios asociados al rol Enriquecedor de Consulta Normalizada

Servicio	Entradas	Salidas	Pre-condiciones	Post-condiciones
Solicitud Enriquecer Consulta				
Resultado Consulta	Enriquecer idUsuario consultaNormalizada	idUsuario consultaEnriquecida	idUsuario ≠ null consultaNormalizada ≠ null	consultaEnriquecida ≠ null

**Tabla 5.** Modelo de Servicios asociado al rol Enriquecedor de Consulta Normalizada.

En la tabla 6., se describe el Modelo de Servicios asociados al rol Ejecutor de Consulta Índice

Servicio	Entradas	Salidas	Pre-condiciones	Post-condiciones
Solicitud Ejecutar Consulta				
Entregar Resultado Ejecucion	idUsuario consultaEnriquecida	idUsuario resultadoConsulta	idUsuario ≠ null consultaEnriquecida ≠ null	resultadoConsulta ≠ null

**Tabla 6.** Modelo de Servicios asociado al rol Ejecutor de Consulta.

En la tabla 7., se describe el Modelo de Servicios asociados al rol Ejecutor de Consulta al Motor.

Servicio	Entradas	Salidas	Pre-condiciones	Post-condiciones
Solicitud Ejecutar Consulta				
Entregar Resultado Consulta	consultaEnriquecida	resultadoConsulta	consultaEnriquecida ≠ null	resultadoConsulta ≠ null

**Tabla 7.** Modelo de Servicios asociado al rol Ejecutor de Consulta al Motor.

En la tabla 8., se describe el Modelo de Servicios asociados al rol Ordenador de Resultados

Servicio	Entradas	Salidas	Pre-condiciones	Post-condiciones
Solicitud Ordenamiento				
Entregar Resultado Ordenamiento	idUsuario resultadoValidacion	idUsuario resultadoOrdenacion	idUsuario ≠ null resultadoValidacion ≠ null	resultadoOrdenamiento ≠ null

**Tabla 8.** Modelo de Servicios asociado al rol Ordenador de Resultados.

En la tabla 9., se describe el Modelo de Servicios asociados al rol Validador de Resultados.

Servicio	Entradas	Salidas	Pre-condiciones	Post-condiciones
Solicitud Validacion				
Entregar Resultado Validacion	resultadoConsulta	resultadoValidacion	resultadoConsulta ≠ null	resultadoValidacion ≠ null

**Tabla 9.** Modelo de Servicios asociado al rol Validador de Resultados.

En la tabla 10., se describe el Modelo de Servicios asociados al rol Gestor de Actualización.

Servicio	Entradas	Salidas	Pre-condiciones	Post-condiciones
Solicitud Actualizacion	tiempoActualizacion		tiempoActualizacion >= tiempoMaximoActualizar	tiempoActualizacion=0
Adicionar Nueva Consulta	nuevaConsulta		nuevaConsulta ≠ null	nuevaConsulta registrada
Informar Nueva Consulta		solNuevaConsulta	nuevaConsulta ≠ null	solNuevaConsulta ≠ null

**Tabla 10.** Modelo de Servicios asociado al rol Gestor de Actualización.

En la tabla 11., se describe el Modelo de Servicios asociados al Rol Gestor de Índice.

<b>Servicio</b>	<b>Entradas</b>	<b>Salidas</b>	<b>Pre-condiciones</b>	<b>Post-condiciones</b>
CrearIndice	consultasOntologia		Indice = null	Indice ≠ null
Actualizar Indice	paginasValidadas		paginasValidadas ≠ null	paginasValidadas indexadas

**Tabla 11.** Modelo de Servicios asociado al rol Gestor de Índice.

➤ Modelo de Relaciones

Las figuras 26. y 27., ilustran las relaciones existentes entre los diferentes agentes del sistema.

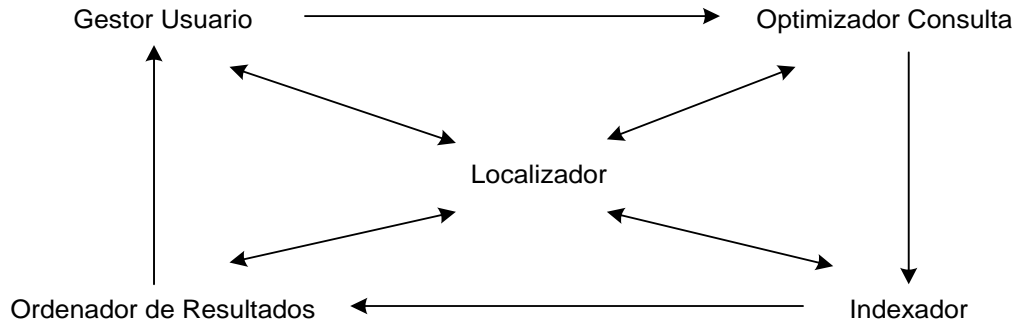


Figura 26. Modelo de relaciones Consultar.

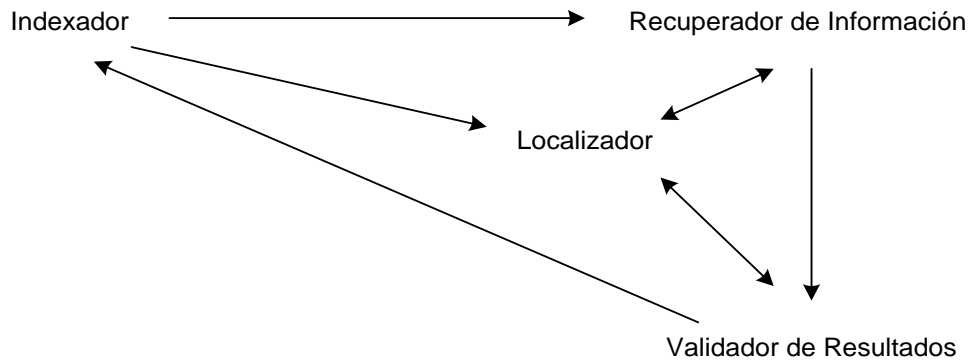


Figura 27. Modelo de relaciones Indexar.

➤ Diagramas de Colaboración:

La figura 28., ilustra el diagrama de colaboración Consultar Recursos en el Sistema.

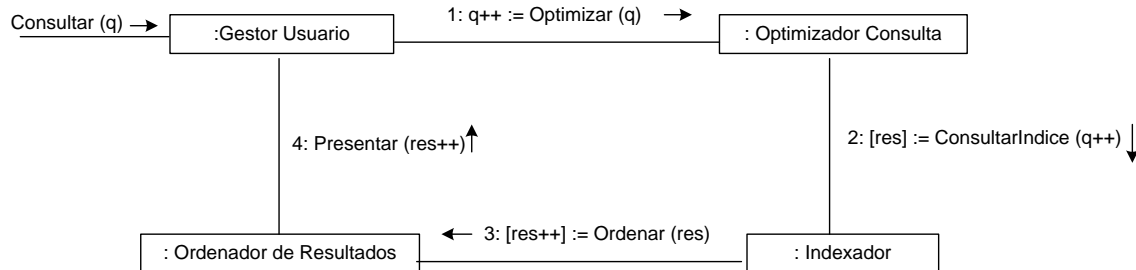


Figura 28. Diagrama de Colaboración Consultar Recursos en el Sistema.

La figura 29., ilustra el diagrama de colaboración Actualizar Índice.

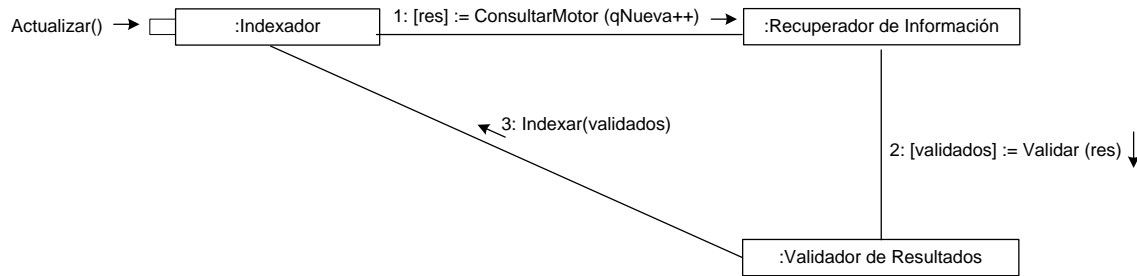


Figura 29. Diagrama de Colaboración Actualizar Índice.

La figura 30., ilustra el diagrama de colaboración Detallado Consultar Motor:

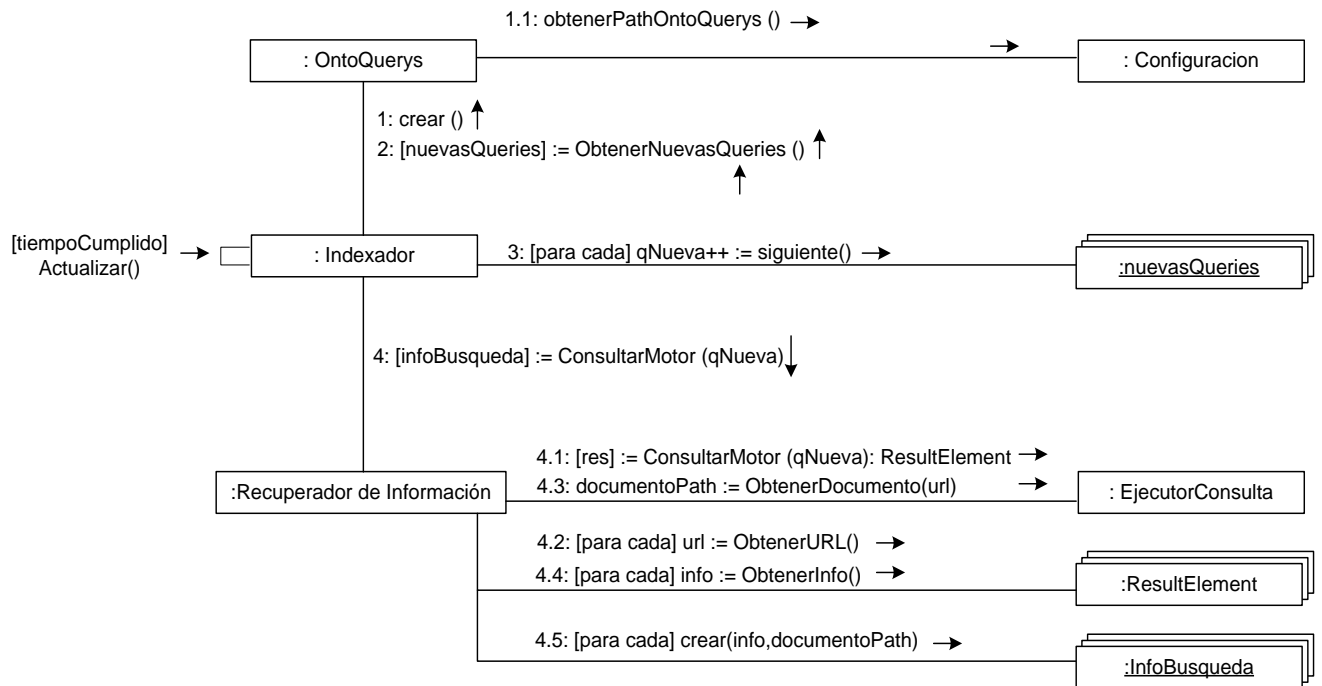


Figura 30. Diagrama de Colaboración Detallado Consultar Motor.



La figura 31., ilustra el diagrama de colaboración Detallado Validar:

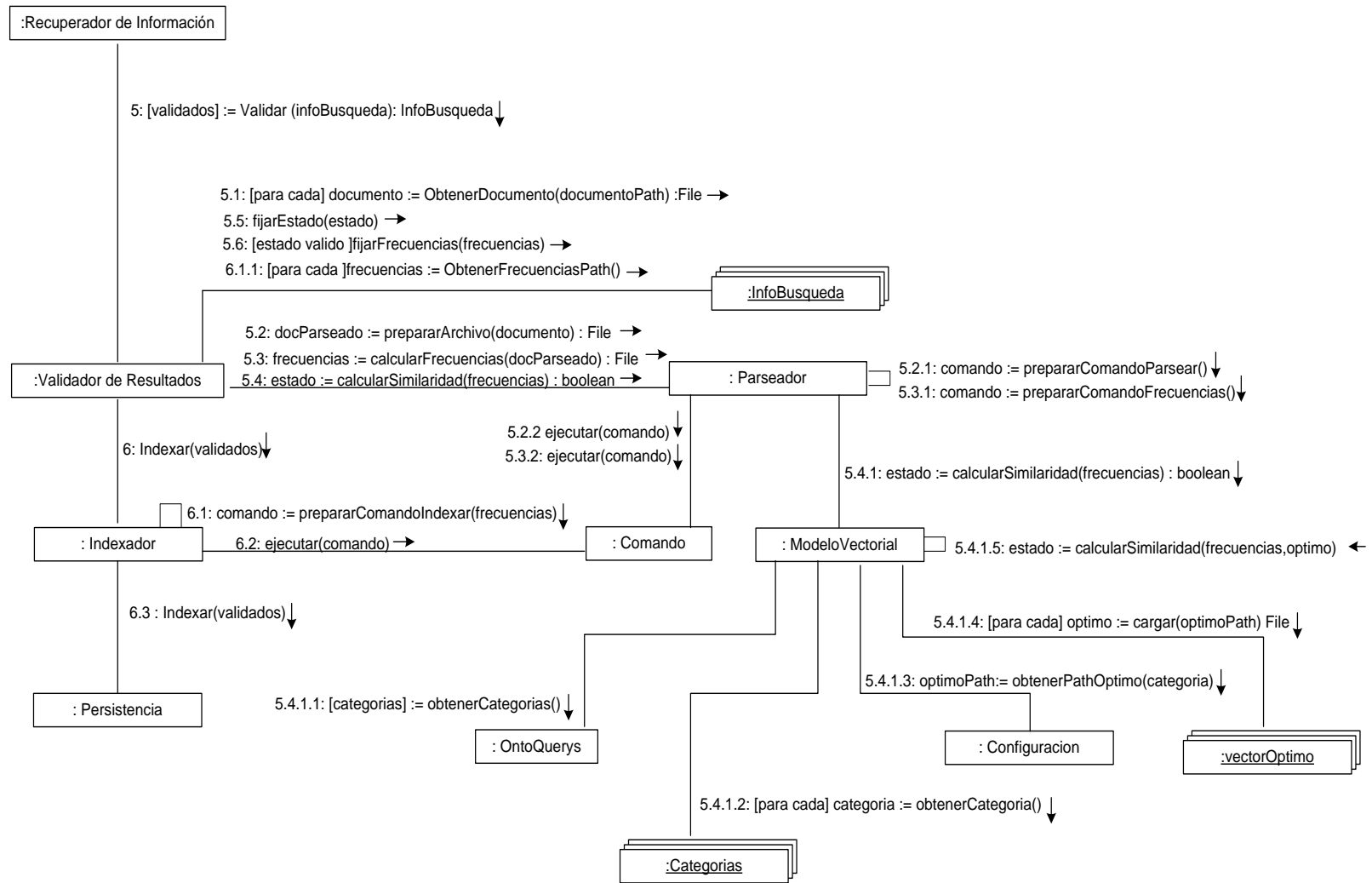


Figura 31. Diagrama de Colaboración Detallado Validar.

### 3.2.4 Implementación

El Sistema Multiagente (MAS) fue implementado con JADE (Java Agent Development Framework) [TIL, 2003] y con la utilización de otras herramientas y tecnologías como Karpanta [FIG, 2003], Interbase 6.0 [WAR, 1985], Protégé Plugin OWL [HOL, 2004] y Google Web APIs service [GOO, 2002], JSP [SDV, 2005a] y el servidor web Tomcat [RUB, 2005], las cuales son descritas con mayor detalle en el capítulo 4.

En adelante se presentará una descripción detallada del Sistema Multiagente., los agentes que lo constituyen, su interacción y sus tareas específicas.

El MAS esta conformado por los siguientes agentes: El agente gestor de Usuario (AGU), El agente gestor de índice (AGI), El Agente de Recursos Web (ARW), El Agente Optimizador de Consulta (AOC), El agente de validación (AV),

- El agente gestor de Usuario (AGU): Este agente es quien presenta al usuario la interfaz de búsqueda y una interfaz que despliega los resultados de la consulta. Para darle funcionalidad a este agente, se creo una pagina JSP, que existe en un contenedor provisto por el servidor web Tomcat [WAR, 1985] y por medio de la cual se logra la interacción con el usuario. La página JSP tiene un AGU como atributo de aplicación, el cual, es registrado en la plataforma. Cuando en el JSP se recibe una petición del usuario, este la notifica al AGU, el cual envía un mensaje tipo REQUEST al agente encargado de retornar la colección de referencias a documentos que resuelvan satisfactoriamente la solicitud de búsqueda hecha. El AGU posee un comportamiento cíclico de escucha para detectar respuestas provenientes de otro agente. Al notificar el JSP la petición, el AUI inicia un proceso bloqueante que culmina solo hasta cuando se tiene una respuesta que dar al usuario, de ahí que el tiempo que tardan en ejecutarse las acciones de los agentes para dar respuesta a esta solicitud, se convierta en un factor crítico para la usabilidad del sistema.
- El agente gestor de índice (AGI): se encarga de mantener actualizado el índice, mediante un proceso de actualización periódico para el cual es programado, recibe las páginas que han sido validadas, verifica si las páginas que recibe están en el índice y adiciona aquellas que no se encuentran indexadas.

Este agente posee tres comportamientos a través de los cuales cumple las responsabilidades antes mencionadas:

- El primer comportamiento es el de actualización del índice, cada cierto periodo de tiempo (inicialmente un mes), el AGI inicializa el proceso de actualización solicitando al Agente Optimizador de Consultas recuperar las consultas a ser ejecutadas en el Motor de Búsqueda, con el fin de extraer los recursos Web a ser validados e indexados.
- El segundo comportamiento es el de Indexar Recursos, comportamiento cíclico que recibe del Agente de Validación la solicitud de indexar un grupo de documentos que han sido validados y adiciona al índice aquellos que no se encuentren en él.
- El tercer comportamiento es el de Consultar Recursos, mediante el cual recibe del Agente Gestor de Usuario una solicitud de consulta y recupera del índice aquellos recursos que satisfacen los requerimientos de consulta.

- El Agente de Recursos Web (ARW): es quien ejecuta las consultas optimizadas usando el Google Web APIs service de google [GOO, 2002] y entrega el resultado al agente de validación. El comportamiento de este agente está a la espera de recibir mensajes con consultas optimizadas para ser ejecutadas sobre el motor de búsqueda que, en este caso, es el google APIs Service [GOO, 2002]. Como google no permite hacer uso de su servicio de búsqueda por simple invocación de la url, ha puesto al servicio de los desarrolladores un API que, aunque restringe el número de consultas al día (mil como máximo), si permite el acceso a tres servicios: búsqueda, spelling y cache.

Utilizando este servicio, el ARW recupera del índice de Google, grupos de 10 recursos y extrae de cada resultado la información relevante para el sistema: título, URL, tamaño en caché. Posteriormente, se descarga el documento para su análisis por parte del agente de validación.

- El Agente Optimizador de Consulta (AOC): es el encargado de seleccionar las consultas que serán enviadas al Agente de Recursos Web, de manera que las consultas que se hagan sobre el motor de búsqueda sean adecuadas en el dominio modelado.

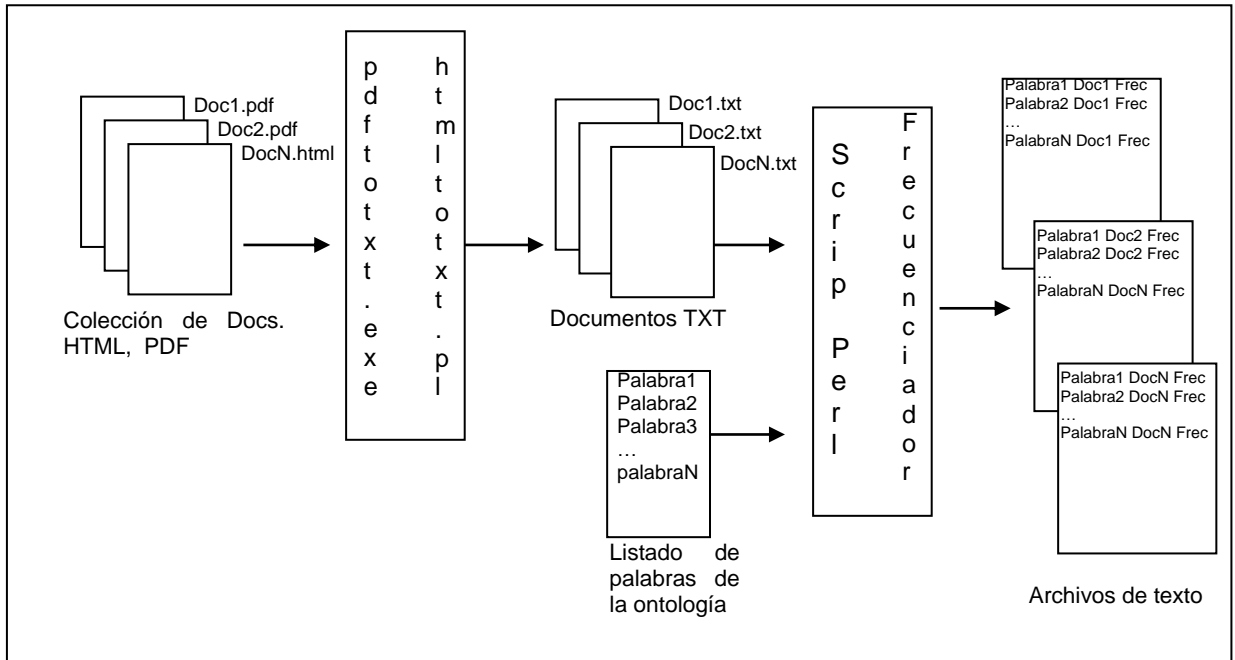
- El agente de validación (AV) El Agente Optimizador de Consulta (AOC): es el encargado de recibir el resultado de la consulta y revisar cuáles de los recursos web recuperados son relevantes o no en el dominio seleccionado, para lo cual realiza una comparación entre dicho recurso y cada uno de los “Vectores Óptimos” de las categorías del dominio, utilizando para ello el modelo Vectorial.

Para poder determinar dicha medida, se parte de los conceptos matemáticos contenidos en el modelo vectorial (ver Capítulo 1). Tomando como base el modelo del dominio, se dividió la temática en trece grandes categorías estudiadas y aprobadas por el experto en el dominio y que, desde su punto de vista, permiten abordar la temática de forma mas concisa y ordenada. Las categorías son Acabados, Acero de Refuerzo, Aditivos, Agregado, Cemento, Mampostería, Mampostería Arquitectónica, Mampostería Estructural, Mampostería no Estructural, Mortero, Muro, Refuerzo, Unidad de Mampostería. Por cada categoría se realizó una búsqueda manual de documentos válidos; entendiendo por documento válido aquel que contiene información específica de cada categoría. Esta documentación también fue revisada por el experto en el dominio.

La colección de documentos válidos de cada categoría son procesados previamente para poder obtener de ellos información en un formato uniforme.

La primera parte del análisis consiste en encontrar la frecuencia con la que un grupo de palabras preestablecido aparece en cada documento, para esto, se determinó previamente que el grupo de palabras debería ser elegido tomando como base nuevamente el modelo del dominio, debido a que, en este modelo se encuentran a nuestro parecer y al parecer del experto, los términos más importantes de la temática mampostería estructural, los cuales se almacenaron para su posterior uso, en un archivo de texto, El proceso para encontrar la frecuencia de términos, utiliza un script para perl [LIZ, 2004] llamado frecuenciador, el cual, requiere dos entradas, la primera, es el grupo de palabras a contar, almacenadas previamente en un archivo de texto plano. La segunda entrada, es el documento a analizar cuyo formato debe ser texto plano, razón por la cual, debe ser procesado previamente utilizando otros scripts que convierten los archivos pdf o html en archivos con formato de texto plano.

La salida del frecuenciador por cada documento analizado, es un archivo de texto que contiene la siguiente información: la palabra seguida del nombre del documento en la que se encontró y por último la frecuencia de esa palabra en ese documento. La figura 32., ilustra el formato de salida del frecuenciador.

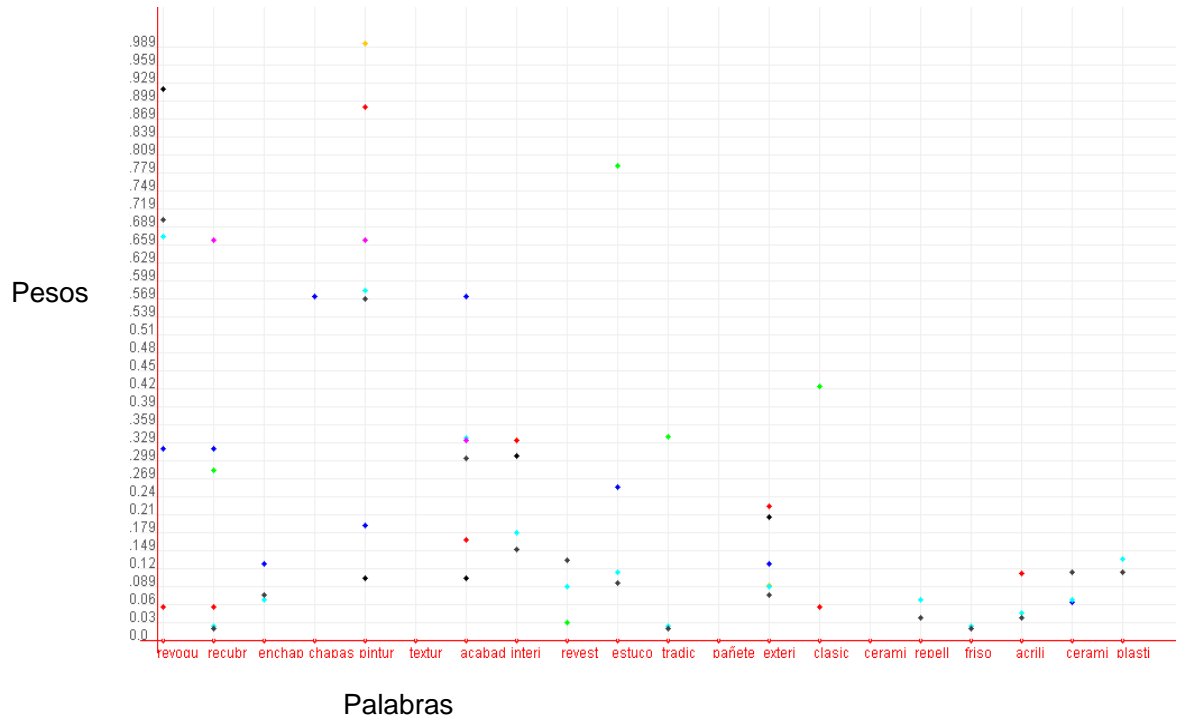


**Figura 32.** Proceso de obtención de frecuencia de palabras en cada documento.

Como se expuso en el capítulo 1, el modelo usado para las operaciones de recuperación y filtrado de información del sistema, es el ampliamente conocido Modelo Vectorial [ZAZ, 2002], el cual intenta recoger la relación de cada documento de una colección de N documentos, con el conjunto de las n características (listado de palabras de la ontología) de la colección.

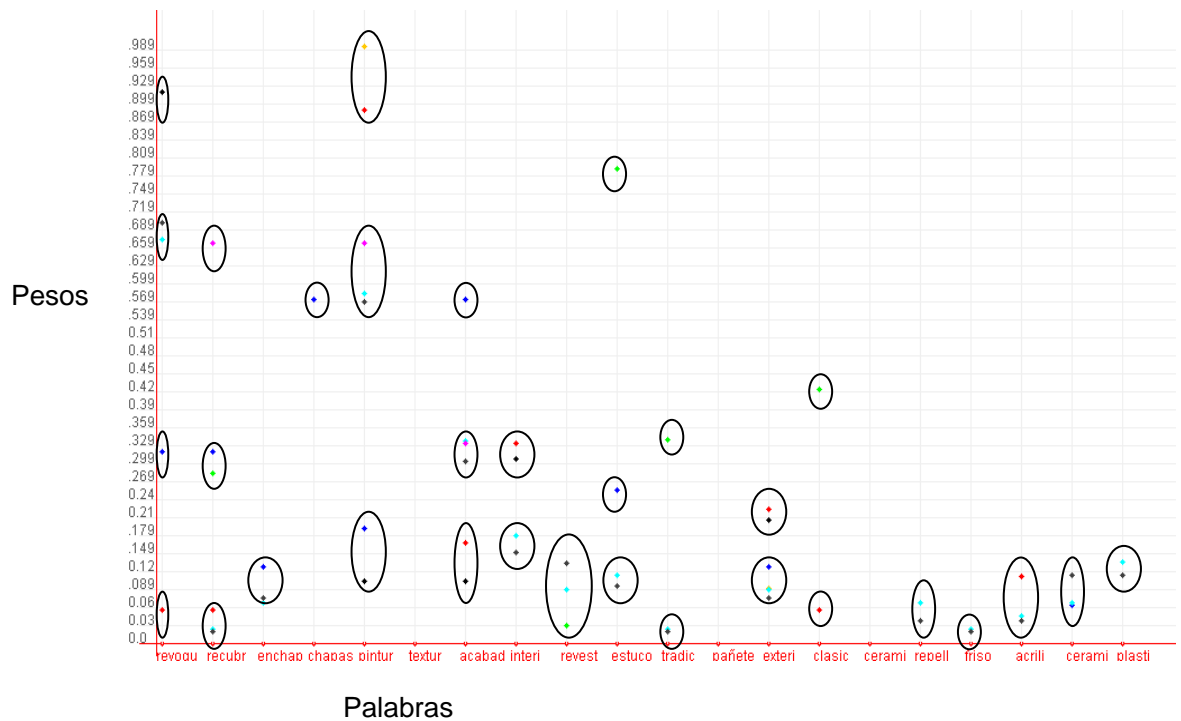
En java se construyó un procedimiento que toma como entrada un archivo con la estructura Termino-Documento-Frecuencia, este procedimiento aplica los conceptos expuestos en este modelo para calcular un vector, que determina en que grado un documento satisface cada una de las n características, es decir, cada elemento de este vector es un valor numérico (peso) que expresa en que grado un documento posee la característica n. Formalmente un documento puede considerarse como un vector que expresa la relación del documento con cada una de esas características.

El procedimiento calcula un vector por cada documento de la colección. Una vez estimados los vectores de cada documento de una categoría, se agruparon sus valores por término y se generó un gráfico de pesos vs. palabras (ver figura 33.).



**Figura. 33.** Grafico Pesos vs. Palabras.

Con base en los resultados obtenidos gráficamente, se pudo apreciar la existencia de valores medios representativos de cada término, en unos casos un solo valor y en otros dos o más de estos valores por término. Para calcular estos valores se fijó una desviación estándar máxima permitida por punto, si un punto esta por fuera del rango permitido, los puntos son reagrupados hasta que todos los puntos queden dentro del valor de la desviación. La figura 34., ilustra este procedimiento.

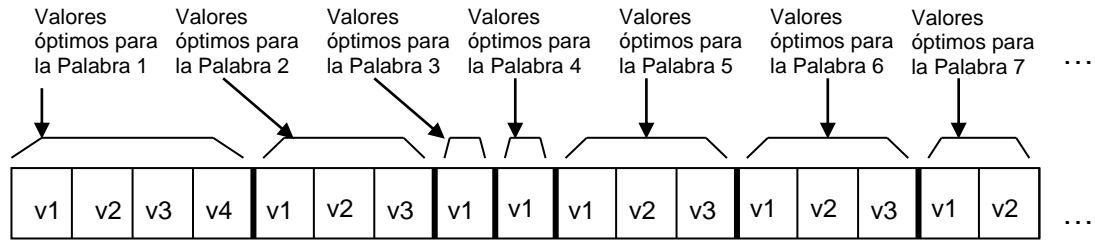


**Figura. 34.** Grafico valores medios representativos de cada término.

La figura 34., muestra 20 palabras representativas de esta categoría y cada palabra con un grupo de puntos que representan los pesos de esas palabras en cada documento de la categoría. Los óvalos representan los grupos que se forman entre los pesos de una palabra y estos grupos son los que finalmente representaran de forma resumida los atributos de esa categoría. Este procedimiento pretende determinar por cada categoría una colección de valores óptimos para cada término.

La colección de términos con sus respectivos valores medios es denominada vector óptimo de la categoría. Como se observa en la figura 34, finalmente se encuentran cuatro valores representativos de la primera palabra de esa categoría, tres valores para la segunda palabra, un valor para la tercera y en adelante como se observa gráficamente. El procedimiento que calcula el vector óptimo, se repitió para las trece categorías y se muestra la estructura final de este vector en la figura 35.

## Categoría 1



**Figura. 35.** Estructura final Vector óptimo representativo de una categoría.

Debido a las características n-dimensionales de los vectores óptimos calculados, es decir, el hecho que cada palabra tenga un número variable de valores representativos, hizo necesario el uso de estructuras flexibles que permitieran su fácil y mejor manipulación, por esta razón se optó por cargarlos sobre tablas de Hash [LOP, 2000].

Para el cálculo del grado de similitud entre dos vectores, es necesario que éstos tengan la misma estructura, es decir, deben ser vectores de una sola dimensión, por lo tanto inicialmente se debe comparar el vector del documento a validar con cada uno de los vectores óptimos de las categorías para obtener vectores óptimos de una sola dimensión. El AV determina cual de los valores representativos de una palabra en el vector óptimo es el más cercano al valor de esa misma palabra en el vector del documento y de esta manera se construye un nuevo vector óptimo pero a diferencia del original, este es de una sola dimensión. La similitud se calcula entre cada uno de los nuevos vectores unidimensionales óptimos encontrados para cada categoría y el vector del documento a validar.

AV recupera los vectores óptimos de cada categoría y recibe un vector de pesos representativo del documento a validar. Determina cuales son los valores más cercanos entre los vectores óptimos y el vector que representa el documento a validar ver Figura. 35.

Se calcula la similitud entre los vectores óptimos y el vector a validar, usando la formula:



$$SIMIL(Q, D_i) = \frac{\sum_{j=1}^m p_j d_{ij}}{\sqrt{\sum_{j=1}^m p_j^2 \cdot \sum_{j=1}^m d_{ij}^2}} \quad (\text{Ver capítulo 1})$$

Se determina cual de las distancias calculadas es la menor y por último se verifica si esta distancia es lo suficientemente pequeña para considerar el documento como válido.

Si el documento es válido se notifica al AGI para que lo indexe.

- El agente Timer (AT): se le adicionó un comportamiento que programa una tarea mensual consistente en enviar al AE los valores obtenidos del atributo query de la ontología modificada, para que sean ejecutados en el motor de búsqueda.

El AGI crea un índice temporal con los resultados de las ejecuciones realizadas por el AE a petición del AT, y se efectúa la actualización del índice reemplazando el índice existente con el temporal creado.

## 4. HERRAMIENTAS TECNOLÓGICAS EMPLEADAS PARA LA IMPLEMENTACIÓN DEL SISTEMA

### 4.1 TECNOLOGIA JAVA.

En el proceso de desarrollo para el Sistema multiagente de optimización de búsquedas en el dominio mampostería estructural basado en técnicas de web semántica se identificaron tres necesidades: (1) la implementación del modelo de dominio en la temática mampostería estructural (ver capítulo 2) (2) la implementación del sistema multiagente y (3) la implementación del prototipo de portal web por medio del cual será puesto a prueba el sistema multiagente previa integración con el modelo del dominio.

La tecnología escogida para la implementación del sistema multiagente fue Java [JAV, 2005], tecnología constituida por el lenguaje de programación Java y por una selección de plataformas especializadas, la plataforma utilizada fue Java 2 SDK, Standard Edition (J2SE) versión 1.4, plataforma que provee un ambiente completo para el desarrollo de aplicaciones de escritorio y servidores. Se eligió la tecnología Java, debido a la experiencia que ya se tenía en el diseño y la implementación de aplicaciones con ella, por

la existencia de APIs (Application Program Interfaces), necesarias para el desarrollo de la aplicación y por ser una tecnología libre para descargar y usar en programación académica y comercial.

## **4.2 PLATAFORMA PARA LA IMPLEMENTACION DE SISTEMAS MULTIAGENTE.**

### **4.2.1 JADE (Java Agent DEvelopment Framework).**

JADE (Java Agent DEvelopment Framework) [TIL, 2003], es un framework implementado completamente en lenguaje Java. Jade simplifica la implementacion de sistemas multiagente a través de un middle-ware que cumple con las especificaciones FiPA y a través de un grupo de herramientas gráficas que soportan las fases de desarrollo y pruebas.(debugging). La plataforma de agentes puede ser distribuida a través de máquinas, las cuales incluso, no necesitan compartir el mismo sistema operativo y la configuración puede ser controlada vía interfaz remota. La configuración puede ser incluso cambiada en tiempo de ejecución por el traslado de los agentes de una máquina a otra, como y cuando sea requerido. JADE esta completamente implementado en lenguaje Java y el requerimiento mínimo del sistema es la versión JDK 1.4 de Java.

JADE es un software libre y es distribuido por TiLab, como software de código abierto bajo los términos de la licencia LGPL ( Lesser General Public License Version 2). Desde Mayo de 2003, a un JADE Board ha sido creado para supervisión y manejo de los proyectos JADE. Actualmente el JADE Board esta conformado por 5 miembros: TiLab, Motorola, Whitestein Technologies AG, Profactor GmbH y France Telecom R&D. La última versión de JADE es JADE 3.3 liberada el 2 de Mayo de 2005 [TIL, 2003]

## **4.3 TECNOLOGIA PARA LA MANIPULACION DE ONTOLOGIAS.**

### **4.3.1 Protégé OWL Plugin.**

El Protégé OWL API es una librería Java open source para OWL y RDF(S) [HOL, 2004]. El API provee clases y métodos para cargar y guardar archivos OWL, consultar y manipular modelos de datos OWL y ejecutar razonamientos. Además, el API esta optimizado para la implementación de interfaces gráficas de usuario. Las versiones anteriores al Protégé OWL API 1.3 constituían básicamente una capa extra sobre el API

core, mientras que las versiones mas recientes, han sido completamente rediseñados para proveer una vista OWL/RDF detallada, que esconde completamente los detalles de implementación subyacentes.

## **4.4 HERRAMIENTA PARA LA EJECUCION DE CONSULTAS EN INTERNET**

### **4.4.1 Google APIs Service.**

El sistema google ofrece un conjunto de librerías para utilizar parte de sus servicios de forma restringida a través del google api. Los servicios que ofrece el sistema google se presentan a continuación:

- **Búsqueda:** el servicio de búsqueda provisto por el google APIs Service recibe la consulta, la ejecuta y entrega los resultados en grupos de máximo 10 resultados. Estos resultados son estructurados de forma que es posible extraer información tal como la url, si esta disponible en cache, tamaño del archivo, ruta del recurso sobre el índice de google, título, y resumen del documento
- **Spelling:** dada una consulta, este servicio entrega sugerencias de corrección a errores gramaticales si los hay, además de una adecuación de la consulta según criterios de google.
- **Cache:** Por medio de este servicio, es posible acceder a todas las url almacenadas en el cache de google, el formato de los archivos del cache es únicamente html aun cuando el recurso original tenga formato pdf, doc, ppt u otro.

## **4.5 SERVIDOR DE PÁGINAS WEB**

### **4.5.1 Tomcat**

Apache Tomcat es un servidor web que funciona como un container de servlets [SDV, 2005b] y JSP [SDV, 2005a]. Su desarrollo se enmarca dentro del proyecto de código abierto Yakarta [ASF, 1999] de Apache y liberado bajo las licencias del Apache Software [ASF, 2004]. Su container implementa las especificaciones de servlets y Java Server Pages (JSP) definidas por la Comunidad de Desarrollo de Especificaciones en la Tecnología Java [JCP, 2003].

Dado que Tomcat fue escrito en Java, hereda su definición multiplataforma. Tomcat surgió de Sun Microsystems cuando desarrollaban un “servidor de páginas” que utilizara Java y posteriormente cedieron el código fuente a la Apache Software Foundation. Contempla una gran cantidad de funcionalidades como: threading, manutención de sesiones, conectividad con el servidor de páginas, es por esto que al servidor web también se le denomina Web-Container [RUB, 2005]. Fácil de extender en funcionalidades como, por ejemplo, al poder integrar Xalan para parsear páginas XML o al agregar James y poder enviar mails desde el servidor, etc.

En la actualidad, Tomcat es un servidor ampliamente conocido y es utilizado con servidores de páginas. Tomcat se entrega bajo la licencia de software para Apache y es open source [RUB, 2005].

El servidor web es el encargado de crear un puente de comunicación entre los usuarios, con sus peticiones, y la aplicación web. En el caso particular de las aplicaciones escritas en Java, el servidor web deja bajo el control del container la ejecución de las clases servlets y le provee la conexión a los flujos asociados a la petición del usuario. Para los JSPs la primera vez de su invocación se transforman en clases servlets y reciben un tratamiento idéntico al determinado para estos.

El servidor web, requiere de un JDK (Java Development Kit) el cual llevará a cabo la ejecución de los programas Java.

## **5 PROTOTIPO DE PORTAL WEB**

El prototipo de portal web desarrollado para probar el sistema multiagente de filtrado, fue diseñado teniendo en cuenta aspectos importantes del área Interacción Humano Computador (HCI: Human Computer Interaction) como es la usabilidad.

En 1979 se dieron las primeras definiciones del término usabilidad en un intento por describir la efectividad del desempeño humano [BEN, 1979]. Una de las definiciones más prácticas de usabilidad la da [KRU, 2000]: “Después de todo, la usabilidad realmente significa estar seguro de que algo funciona bien: que una persona con habilidades promedio (e incluso por debajo del promedio) pueda utilizar una cosa - ya sea un sitio web, un jet de combate, o una puerta rotatoria - para su intencionado sin terminar

enormemente frustrado". Y más concretamente, la norma ISO 9241 (Ergonomic Requirements for Visual Display Terminals. [ISO, 2001] define la usabilidad como "la facilidad de uso de una aplicación informática".

La usabilidad no es solo una cuestión de tamaño, color y formato. Es disponer como verdaderas herramientas a las aplicaciones y ello no se consigue con el simple embellecimiento de una interfaz de usuario [BAE, 2003].

La idea principal a seguir en el diseño del prototipo de portal web para esta aplicación es proveer al usuario una interfaz simple pero funcional con la cual se sienta cómodo y dentro de la cual encuentre los elementos necesarios que satisfagan sus necesidades de búsqueda

Teniendo en cuenta que portal es un término sinónimo de puente, para referirse a un sitio web que sirve o pretende servir como un sitio principal de partida para las personas que se conectan al World Wide Web. Son sitios que los usuarios tienden a visitar como sitios ancla [MIL, 2003]. Los portales tienen gran reconocimiento en Internet por el poder de influencia que tienen sobre grandes comunidades.

La idea es emplear este portal para localizar la información y los sitios de interés del usuario y desde aquí facilitar su actividad en Internet.

## **6 PRUEBAS REALIZADAS**

En este capítulo se presentan las pruebas realizadas al sistema de filtrado de información y el análisis de los resultados obtenidos. Las pruebas fueron realizadas y preparadas con el Ingeniero Luís Fernando Polanco, quien es experto conocedor de la temática y quien verificó la relevancia de los resultados obtenidos.

### **6.1 PRUEBA 1**

Titulo de la prueba: Efectividad absoluta del motor de búsqueda Google [GOO, 2004]

Objetivo:

Determinar el número de documentos realmente válidos en cada una de las trece categorías del modelo de dominio recuperadas por el motor de búsqueda google dada una consulta. Teniendo muy presente la fecha y hora de la ejecución de la prueba

Estrategia de ejecución:

Se construye una consulta relevante en el dominio mampostería estructural específicamente en la categoría a evaluar.

Se ejecuta la consulta por medio del motor de búsqueda google.

Tener en cuenta la fecha y hora de la ejecución de la prueba, puesto que la información indexada en google permanece sometida a cambios constantes

Determinar cuantos documentos de los recuperados con la consulta van a ser analizados manualmente. Esta observación es importante puesto que en la mayoría de los casos google recupera cientos de miles de páginas por consulta los cuales son grupos difícilmente analizables manualmente.

Determinar en lotes de cuantos documentos se realizará el análisis manual. Los resultados de la prueba se presentarán agrupados dependiendo de esta división por lotes.

Se realizan tantas consultas como categorías tiene el dominio.

Se ejecuta la consulta sobre el motor de google y se inicia el análisis manual de un número de documentos previamente determinado, registrando cuantos documentos relevantes en la categoría se encuentran por cada lote de documentos.

Escenario:

Datos de entrada: consultas previamente definidas para cada categoría.

Motor de búsqueda utilizado para la prueba: Google.

Número de categorías evaluadas: 13

Número de consultas: una por categoría

Caso 1: Categoría Acabados

Consulta: acabados para mampostería

Fecha y hora de realización de la prueba: Noviembre 4 de 2005. 9:31 a.m.

Documentos a ser analizados manualmente: 260 en lotes de 10 documentos con los primeros 40 documentos, de ahí en adelante lotes de 20 documentos hasta los 100 primeros documentos y finalmente lotes de 40 documentos hasta completar los 260 documentos. Se determinó con base en la experiencia de usuario que los documentos mas relevantes son recuperados dentro de los primeros 40 resultados arrojados por el motor google.

Resultados obtenidos:

Recursos recuperados por el motor de búsqueda google: 14600 documentos.

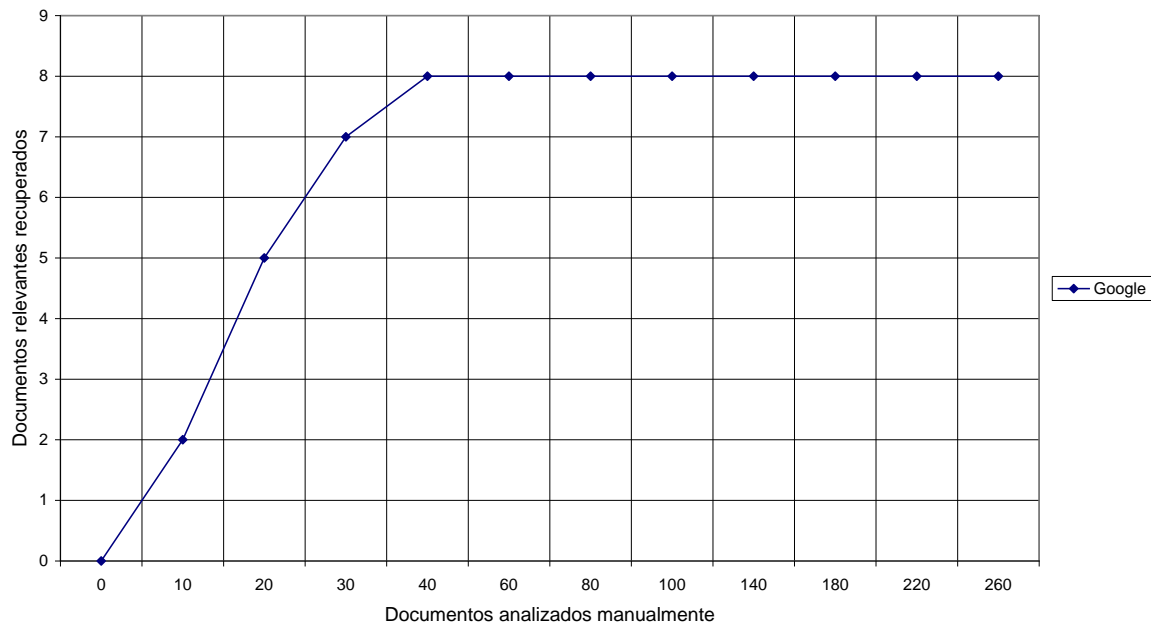
Documentos relevantes recuperados dentro de los 260 documentos recuperados: 6.

Los resultados se resumen en la tabla número 12., y se ilustran en la figura 36.

Documentos analizados manualmente.	10	20	30	40	60	80	100	140	180	220	260
Documentos relevantes recuperados.	2	3	4	6	6	6	6	6	6	6	6

**Tabla 12.** Resumen resultados prueba 1 para la categoría acabados.

Categoría Acabados



**Figura 36.** Resultados prueba 1 para la categoría acabados. Documentos relevantes recuperados vs. Documentos analizados manualmente.

Caso 2: Categoría Acero de refuerzo

Consulta: “acero de refuerzo”

Fecha y hora de realización de la prueba: Noviembre 4 de 2005. 11:19 a.m.

Documentos a ser analizados manualmente: 260 en lotes de 10 documentos con los primeros 40 documentos, de ahí en adelante lotes de 20 documentos hasta los 100 primeros documentos y finalmente lotes de 40 documentos hasta completar los 260 documentos. Se determinó con base en la experiencia de usuario que los documentos

mas relevantes son recuperados dentro de los primeros 40 resultados arrojados por el motor google.

Resultados obtenidos:

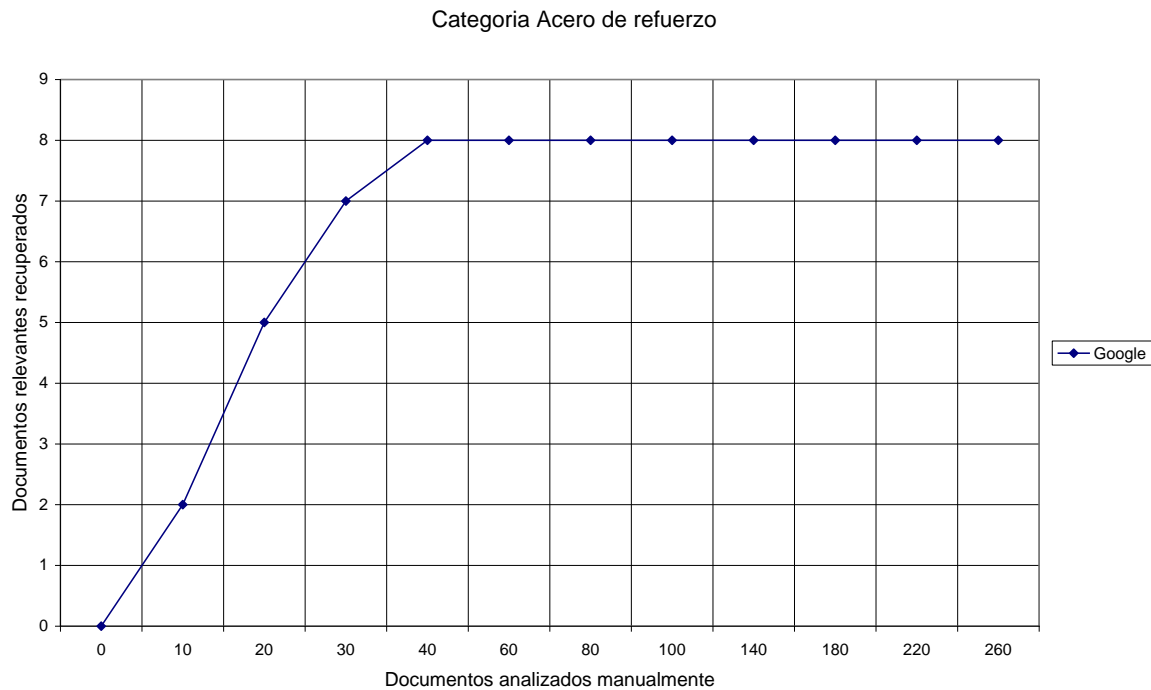
Recursos recuperados por el motor de búsqueda google: 11200 documentos.

Documentos relevantes recuperados dentro de los 260 documentos recuperados: 5.

Los resultados se resumen en la tabla número 13., y se ilustran en la figura 37.

Documentos analizados manualmente.	10	20	30	40	60	80	100	140	180	220	260
Documentos relevantes recuperados.	1	3	5	5	5	5	5	5	5	5	5

**Tabla 13.** Resumen resultados prueba 1 para la categoría Acero de refuerzo.



**Figura 37.** Resultados prueba 1 para la categoría acero de refuerzo. Documentos relevantes recuperados vs. Documentos analizados manualmente.

### Caso 3: Categoría Aditivos

Consulta: aditivos +mortero

Fecha y hora de realización de la prueba: Noviembre 4 de 2005. 2:31 p.m.

Documentos a ser analizados manualmente: 260 en lotes de 10 documentos con los primeros 40 documentos, de ahí en adelante lotes de 20 documentos hasta los 100



primeros documentos y finalmente lotes de 40 documentos hasta completar los 260 documentos. Se determinó con base en la experiencia de usuario que los documentos mas relevantes son recuperados dentro de los primeros 40 resultados arrojados por el motor google.

Resultados obtenidos:

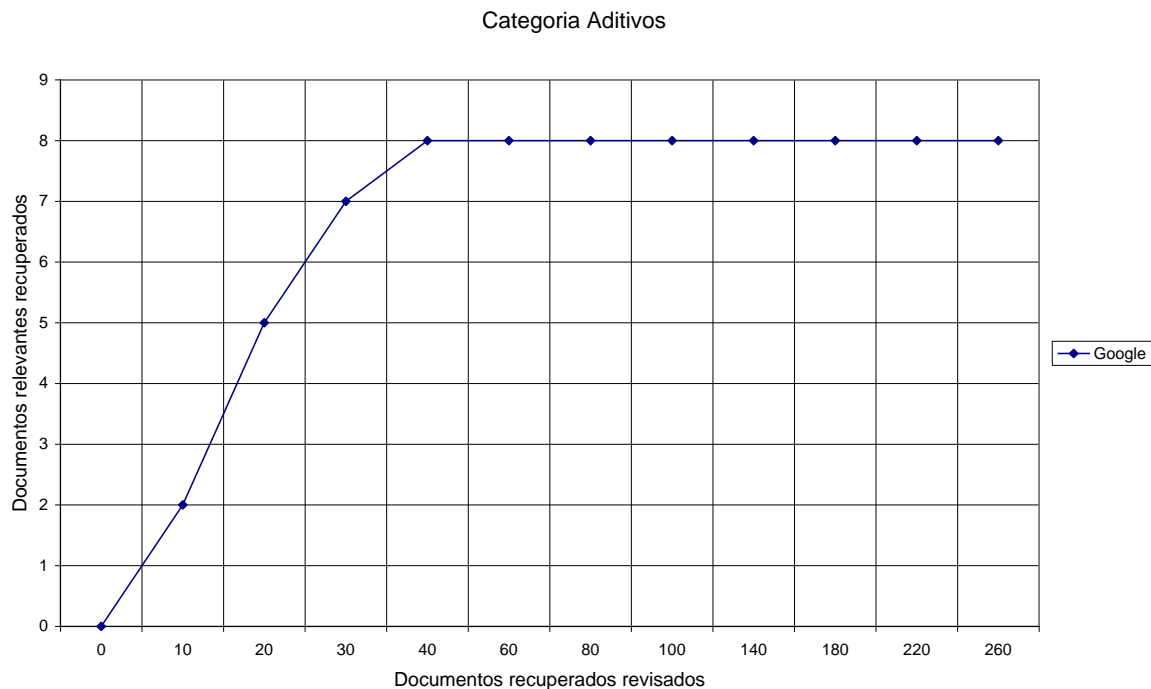
Recursos recuperados por el motor de búsqueda google: 1280 documentos.

Documentos relevantes recuperados dentro de los 260 documentos recuperados: 8.

Los resultados se resumen en la tabla número 14., y se ilustran en la figura 38.

Documentos analizados manualmente.	10	20	30	40	60	80	100	140	180	220	260
Documentos relevantes recuperados.	4	6	6	7	8	8	8	8	8	8	8

**Tabla 14.** Resumen resultados prueba 1 para la categoría aditivos.



**Figura 38.** Resultados prueba 1 para la categoría aditivos. Documentos relevantes recuperados vs. Documentos analizados manualmente.

#### Caso 4: Categoría Agregado

Consulta: tipos de agregados +mortero +mamposteria

Fecha y hora de realización de la prueba: Noviembre 4 de 2005. 4:31 p.m.

Documentos a ser analizados manualmente: 260 en lotes de 10 documentos con los primeros 40 documentos, de ahí en adelante lotes de 20 documentos hasta los 100 primeros documentos y finalmente lotes de 40 documentos hasta completar los 260 documentos. Se determinó con base en la experiencia de usuario que los documentos mas relevantes son recuperados dentro de los primeros 40 resultados arrojados por el motor google.

Resultados obtenidos:

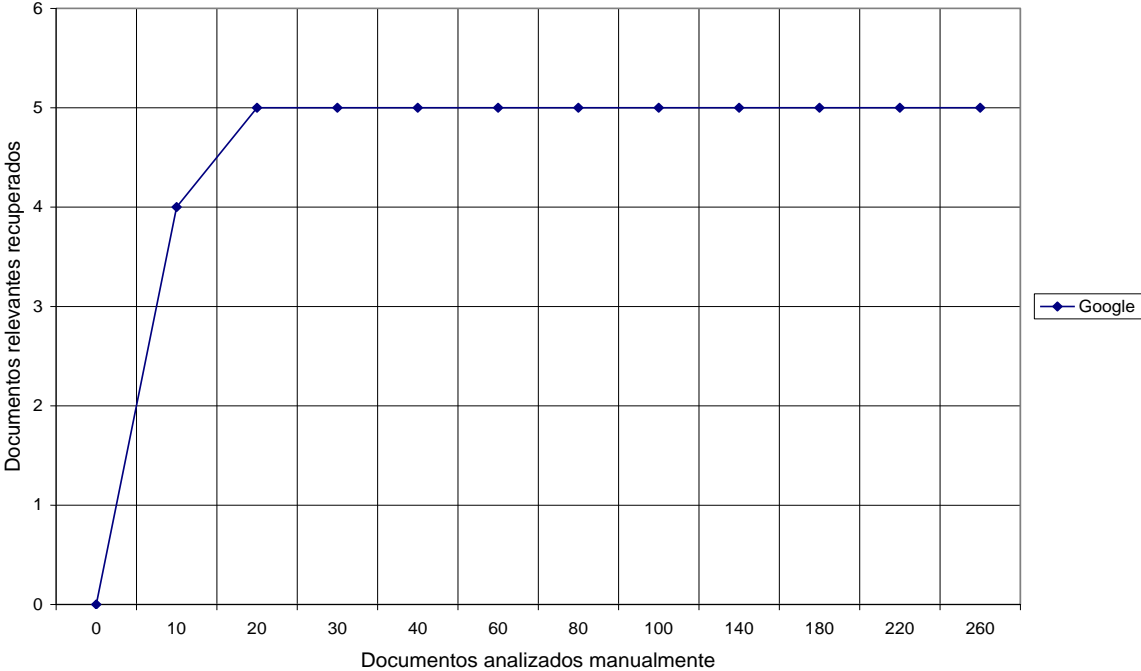
Recursos recuperados por el motor de búsqueda google: 727 documentos.

Documentos relevantes recuperados dentro de los 260 documentos recuperados: 5.

Los resultados se resumen en la tabla número 15., y se ilustran en la figura 39.

Documentos analizados manualmente.	10	20	30	40	60	80	100	140	180	220	260
Documentos relevantes recuperados.	4	5	5	5	5	5	5	5	5	5	5

**Tabla 15.** Resumen resultados prueba 1 para la categoría Agregado.  
Categoría Agregado



**Figura 39.** Resultados prueba 1 para la categoría agregado. Documentos relevantes recuperados vs. Documentos analizados manualmente.

Caso 5: Categoría Cemento

Consulta: "tipos de cemento"

Fecha y hora de realización de la prueba: Noviembre 7 de 2005. 2:35 p.m.

Documentos a ser analizados manualmente: 260 en lotes de 10 documentos con los primeros 40 documentos, de ahí en adelante lotes de 20 documentos hasta los 100 primeros documentos y finalmente lotes de 40 documentos hasta completar los 260 documentos. Se determinó con base en la experiencia de usuario que los documentos mas relevantes son recuperados dentro de los primeros 40 resultados arrojados por el motor google.

Resultados obtenidos:

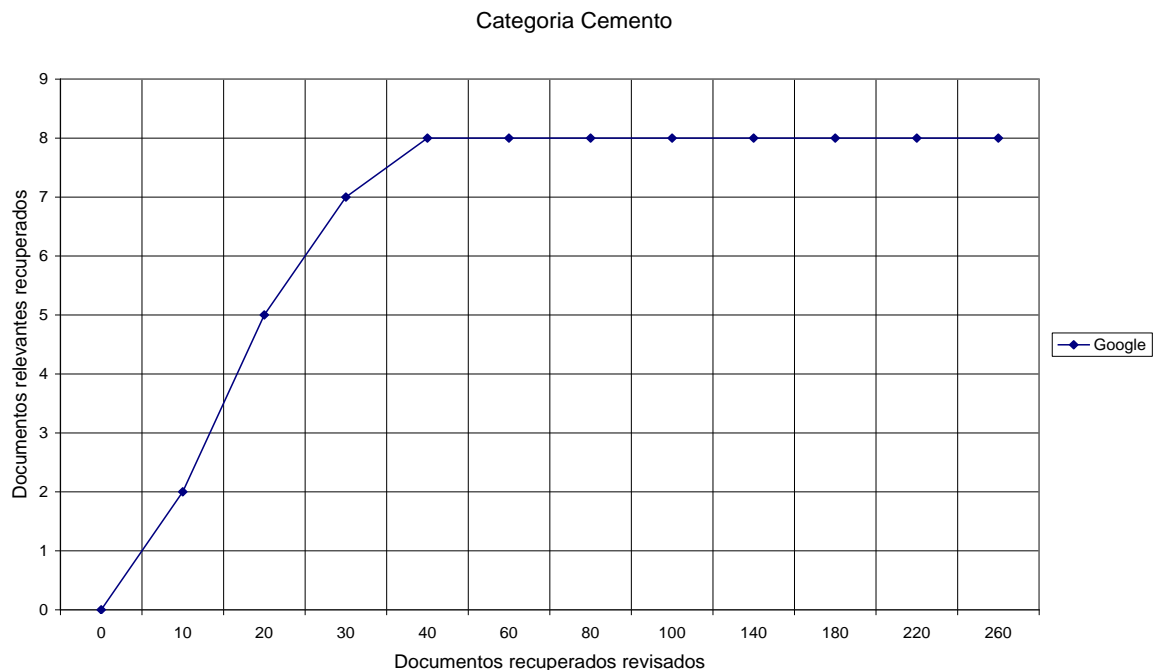
Recursos recuperados por el motor de búsqueda google: 1100 documentos.

Documentos relevantes recuperados dentro de los 260 documentos recuperados: 10.

Los resultados se resumen en la tabla número 16., y se ilustran en la figura 40.

Documentos analizados manualmente.	10	20	30	40	60	80	100	140	180	220	260
Documentos relevantes recuperados.	4	7	8	9	10	10	10	10	10	10	10

**Tabla 16.** Resumen resultados prueba 1 para la categoría cemento.



**Figura 40.** Resultados prueba 1 para la categoría cemento. Documentos relevantes recuperados vs. Documentos analizados manualmente.

Caso 6: Categoría Mampostería

Consulta: "tipos de mampostería"

Fecha y hora de realización de la prueba: Noviembre 7 de 2005. 10:11 a.m.

Documentos a ser analizados manualmente: 151 en lotes de 10 documentos con los primeros 70 documentos, de ahí en adelante lotes de 20 documentos hasta los 151 primeros documentos en este caso en el cual google recupera 151 documentos, se analizaron todos y se divido el análisis manual en mas lotes de 10 documentos (hasta 70 documentos) y en adelante lotes de 20 hasta completar el análisis manual de los 151 documentos

Resultados obtenidos:

Recursos recuperados por el motor de búsqueda google: 151 documentos.

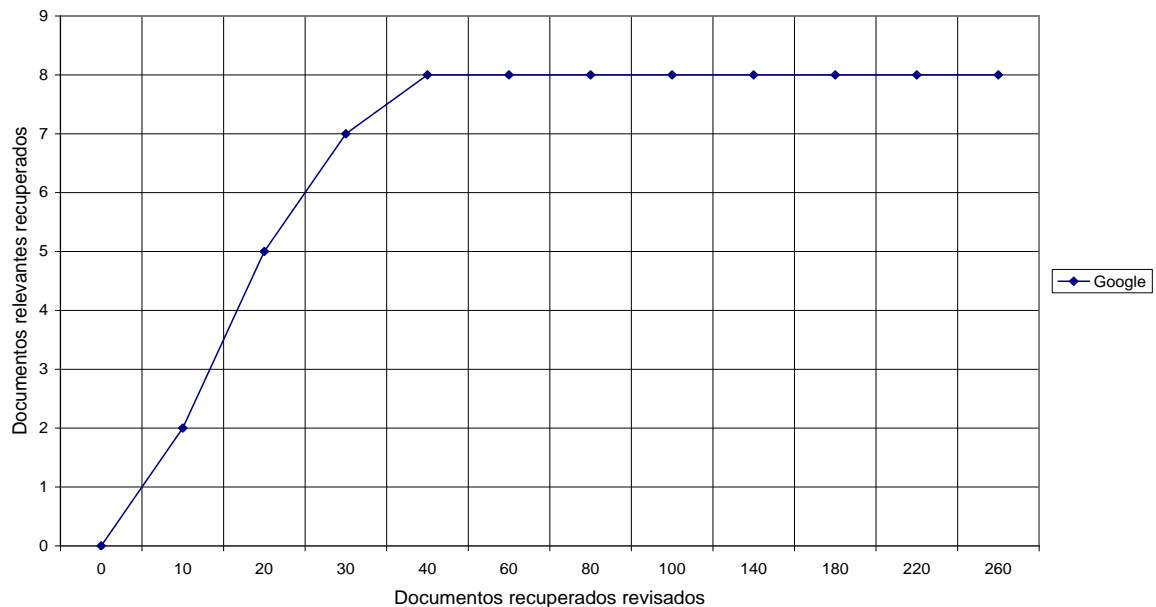
Documentos relevantes recuperados dentro de los 151 documentos recuperados: 8.

Los resultados se resumen en la tabla número 17., y se ilustran en la figura 41.

Documentos analizados manualmente.	10	20	30	40	50	60	70	90	110	130	151
Documentos relevantes recuperados.	3	6	8	8	8	8	8	8	8	8	8

**Tabla 17.** Resumen resultados prueba 1 para la categoría mampostería.

Categoría Mampostería



**Figura 41.** Resultados prueba 1 para la categoría mampostería. Documentos relevantes recuperados vs. Documentos analizados manualmente.

Caso 7: Categoría Mampostería arquitectónica.

Consulta: "tmamposteria architectonica"

Fecha y hora de realización de la prueba: Noviembre 8 de 2005. 10:35 a.m.

Documentos a ser analizados manualmente: 17 en 2 lotes de 10 y 7 respectivamente.

Recursos recuperados por el motor de búsqueda google: 151 documentos.

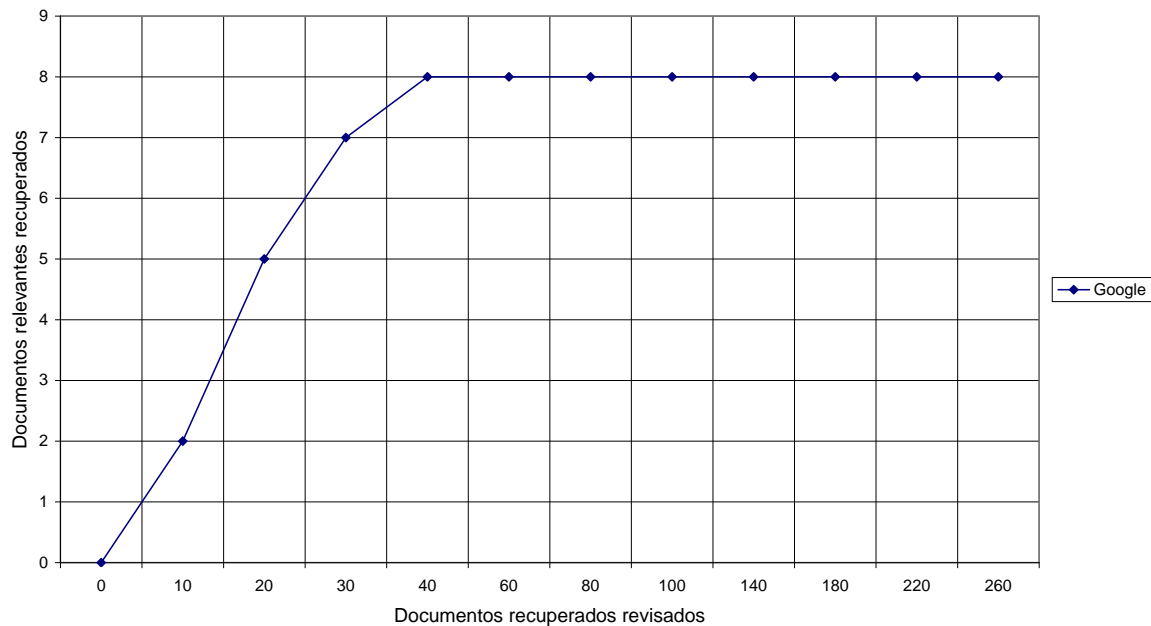
Documentos relevantes recuperados dentro de los 17 documentos recuperados: 3.

Los resultados obtenidos se resumen en la tabla número 18., y se ilustran en la figura 42.

Documentos analizados manualmente.	10	17	-	-	-	-	-	-	-	-	-
Documentos relevantes recuperados.	2	3	-	-	-	-	-	-	-	-	-

**Tabla 18.** Resumen resultados prueba 1 para la categoría mampostería arquitectónica.

Categoría Mamposteria Arquitectonica



**Figura 42.** Resultados prueba 1 para la categoría mampostería arquitectónica. Documentos relevantes recuperados vs. Documentos analizados manualmente.

### Caso 8: Categoría Mampostería estructural

Consulta: "mamposteria estructural"

Fecha y hora de realización de la prueba: Noviembre 8 de 2005. 3:35 p.m.

Documentos a ser analizados manualmente: 260 en lotes de 10 documentos con los primeros 40 documentos, de ahí en adelante lotes de 20 documentos hasta los 100 primeros documentos y finalmente lotes de 40 documentos hasta completar los 260

documentos. Se determinó con base en la experiencia de usuario que los documentos mas relevantes son recuperados dentro de los primeros 40 resultados arrojados por el motor google.

Resultados obtenidos:

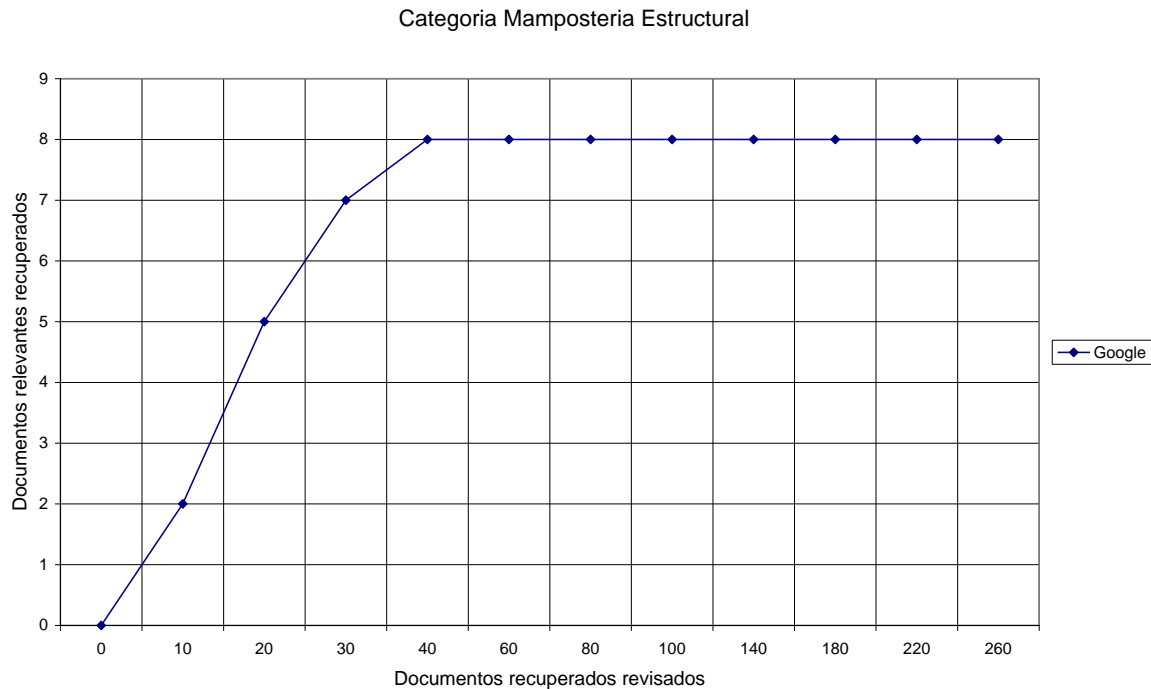
Recursos recuperados por el motor de búsqueda google: 550 documentos.

Documentos relevantes recuperados dentro de los 260 documentos recuperados: 6.

Los resultados se resumen en la tabla número 19., y se ilustran en la figura 43.

Documentos analizados manualmente.	10	20	30	40	60	80	100	140	180	220	260
Documentos relevantes recuperados.	2	3	4	5	6	6	6	6	6	6	6

**Tabla 19.** Resumen resultados prueba 1 para la categoría mampostería estructural.



**Figura 43.** Resultados prueba 1 para la categoría mampostería estructural. Documentos relevantes recuperados vs. Documentos analizados manualmente.

Caso 9: Categoría Mampostería no estructural

Consulta: “mampostería no estructural”

Fecha y hora de realización de la prueba: Noviembre 8 de 2005. 4:39 p.m.

Documentos a ser analizados manualmente: 79 en lotes de 10 documentos con los 70 documentos recuperados, mas relevantes son recuperados dentro de los primeros 40.

Resultados obtenidos:

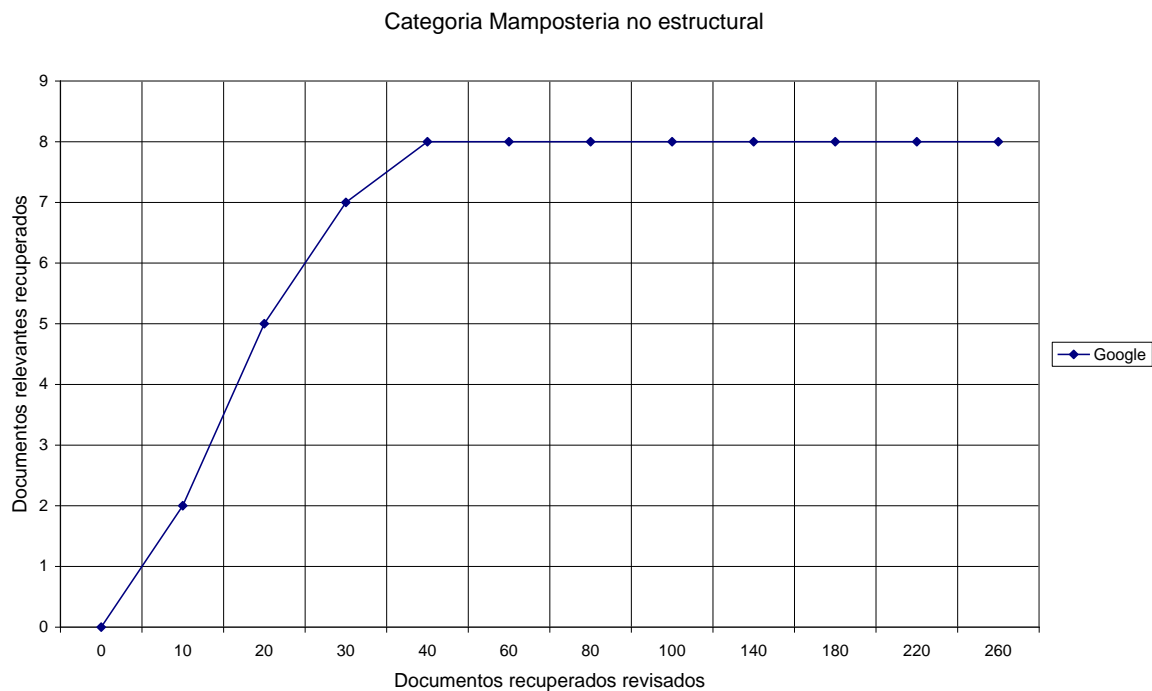
Recursos recuperados por el motor de búsqueda google: 79 documentos.

Documentos relevantes recuperados dentro de los 79 documentos recuperados: 6.

Los resultados se resumen en la tabla número 20., y se ilustran en la figura 44.

Documentos analizados manualmente.	10	20	30	40	50	60	70	80	-	-	-
Documentos relevantes recuperados.	2	4	5	6	6	6	6	6	-	-	-

**Tabla 20.** Resumen resultados prueba 1 para la categoría mampostería no estructural.



**Figura 45.** Resultados prueba 1 para la categoría mampostería no estructural. Documentos relevantes recuperados vs. Documentos analizados manualmente.

Caso 10: Categoría Mortero

Consulta: mortero + mampostería

Fecha y hora de realización de la prueba: Noviembre 9 de 2005. 8:19 p.m.

Documentos a ser analizados manualmente: 260 en lotes de 10 documentos con los primeros 40 documentos, de ahí en adelante lotes de 20 documentos hasta los 100 primeros documentos y finalmente lotes de 40 documentos hasta completar los 260 documentos. Se determinó con base en la experiencia de usuario que los documentos mas relevantes son recuperados dentro de los primeros 40 resultados arrojados por el motor google.

Resultados obtenidos:

Recursos recuperados por el motor de búsqueda google: 884 documentos.

Documentos relevantes recuperados dentro de los 260 documentos recuperados: 8.

Los resultados se resumen en la tabla número 21., y se ilustran en la figura 45.

Documentos analizados manualmente.	10	20	30	40	60	80	100	140	180	220	260
Documentos relevantes recuperados.	3	4	6	8	8	8	8	8	8	8	8

Tabla 21. Resumen resultados prueba 1 para la categoría mortero.

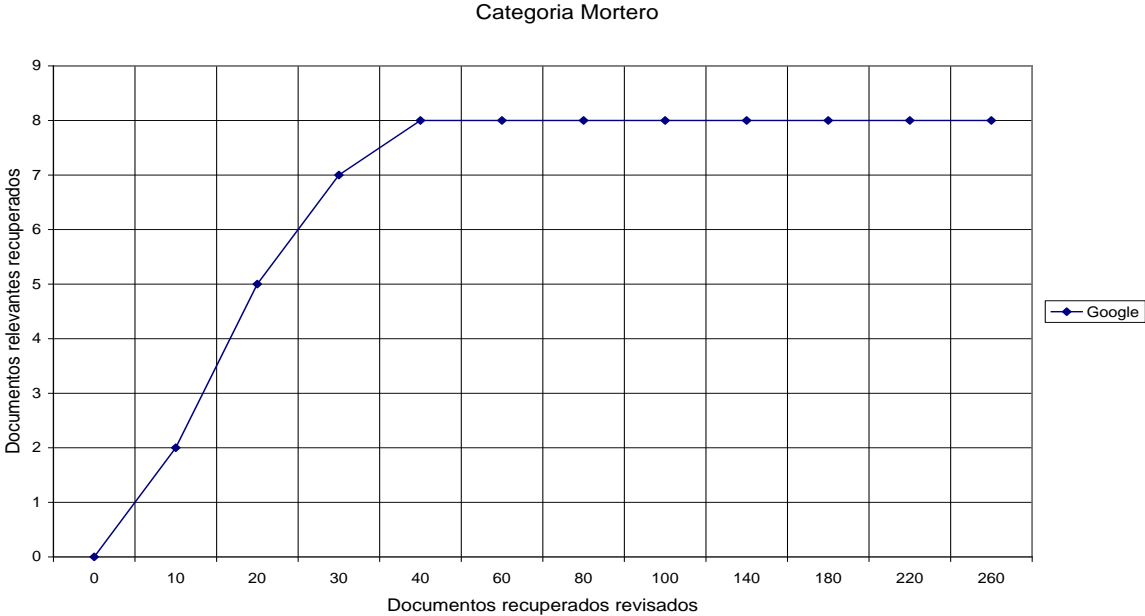


Figura 45. Resultados prueba 1 para la categoría mortero. Documentos relevantes recuperados vs. Documentos analizados manualmente.

Caso 11: Categoría Muro

Consulta: tipología de muros + mampostería



Fecha y hora de realización de la prueba: Noviembre 9 de 2005. 9:56 a.m.

Documentos a ser analizados manualmente: los 262 en lotes de 10 documentos con los primeros 40 documentos, de ahí en adelante lotes de 20 documentos hasta los 100 primeros documentos y finalmente lotes de 40 documentos hasta completar los 262 documentos. Se determinó con base en la experiencia de usuario que los documentos mas relevantes son recuperados dentro de los primeros 40 resultados arrojados por el motor google.

Resultados obtenidos:

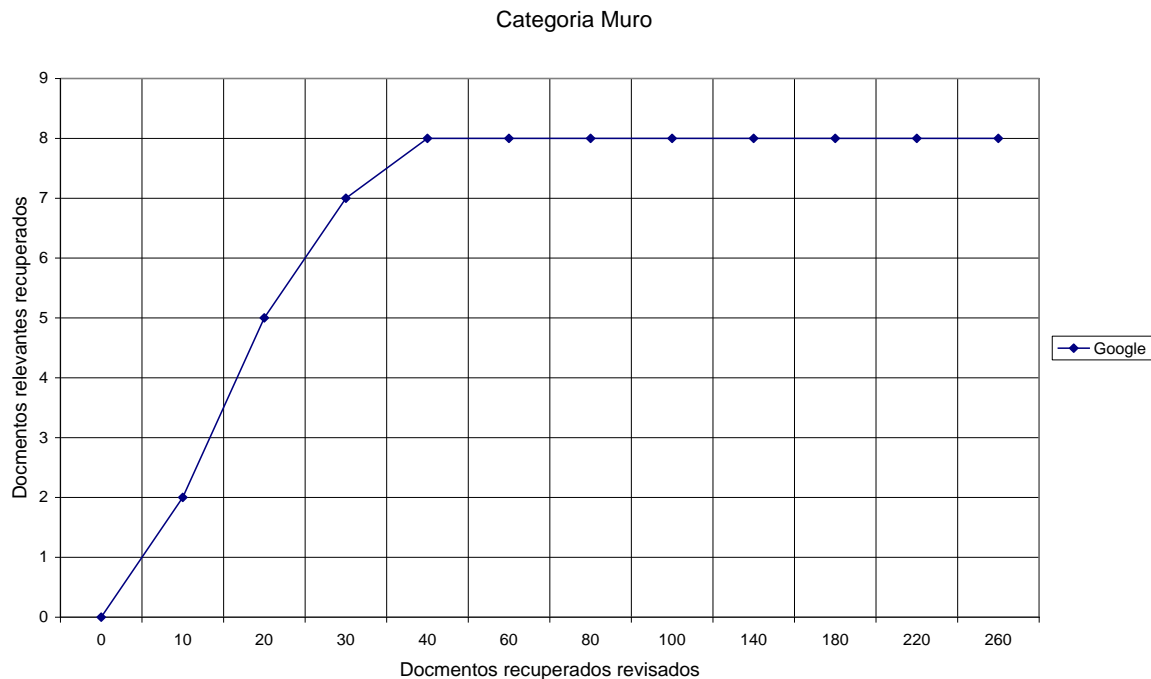
Recursos recuperados por el motor de búsqueda google: 262 documentos.

Documentos relevantes recuperados dentro de los 262 documentos recuperados: 6.

Los resultados se resumen en la tabla número 22., y se ilustran en la figura 46.

Documentos analizados manualmente.	10	20	30	40	60	80	100	140	180	220	262
Documentos relevantes recuperados.	1	3	5	6	6	6	6	6	6	6	6

**Tabla 22.** Resumen resultados prueba 1 para la categoría muro.



**Figura 46.** Resultados prueba 1 para la categoría muro. Documentos relevantes recuperados vs. Documentos analizados manualmente.

## Caso 12: Categoría Refuerzo

Consulta: tipos de refuerzo +mamposteria

Fecha y hora de realización de la prueba: Noviembre 9 de 2005. 11:20 a.m.

Documentos a ser analizados manualmente: los 260 en lotes de 10 documentos con los primeros 40 documentos, de ahí en adelante lotes de 20 documentos hasta los 100 primeros documentos y finalmente lotes de 40 documentos hasta completar los 260 documentos. Se determinó con base en la experiencia de usuario que los documentos mas relevantes son recuperados dentro de los primeros 40 resultados arrojados por el motor google.

Resultados obtenidos:

Recursos recuperados por el motor de búsqueda google: 671 documentos.

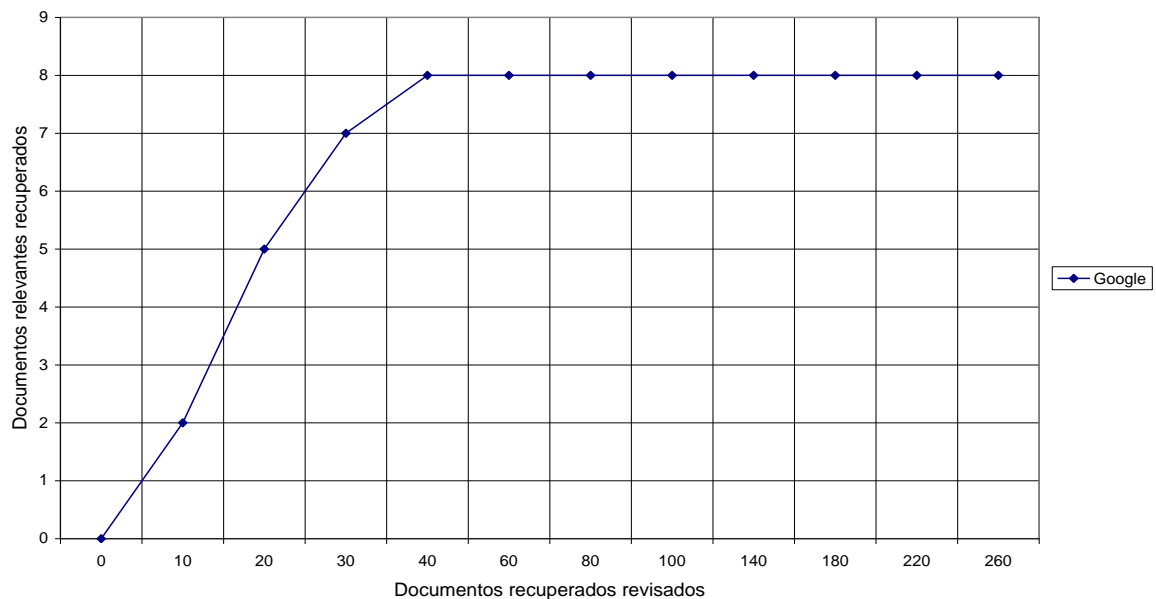
Documentos relevantes recuperados dentro de los 260 documentos recuperados: 5.

Los resultados se resumen en la tabla número 23., y se ilustran en la figura 47.

Documentos analizados manualmente.	10	20	30	40	60	80	100	140	180	220	260
Documentos relevantes recuperados.	2	3	4	5	5	5	5	5	5	5	5

**Tabla 23.** Resumen resultados prueba 1 para la categoría refuerzo.

Categoría Refuerzo



**Figura 47.** Resultados prueba 1 para la categoría refuerzo. Documentos relevantes recuperados vs. Documentos analizados manualmente.

Caso 13: Categoría Unidad de Mampostería

Consulta: unidad de mampostería +tipos

Fecha y hora de realización de la prueba: Noviembre 9 de 2005. 1:20 6.m.

Documentos a ser analizados manualmente: los 260 en lotes de 10 documentos con los primeros 40 documentos, de ahí en adelante lotes de 20 documentos hasta los 100 primeros documentos y finalmente lotes de 40 documentos hasta completar los 260 documentos. Se determinó con base en la experiencia de usuario que los documentos mas relevantes son recuperados dentro de los primeros 40 resultados arrojados por el motor google.

Resultados obtenidos:

Recursos recuperados por el motor de búsqueda google: 27300 documentos.

Documentos relevantes recuperados dentro de los 260 documentos recuperados: 8.

Los resultados se resumen en la tabla número 24., y se ilustran en la figura 48.

Documentos analizados manualmente.	10	20	30	40	60	80	100	140	180	220	260
Documentos relevantes recuperados.	2	5	7	8	8	8	8	8	8	8	8

Tabla 24. Resumen resultados prueba 1 para la categoría mortero.

Categoría Unidad de mampostería

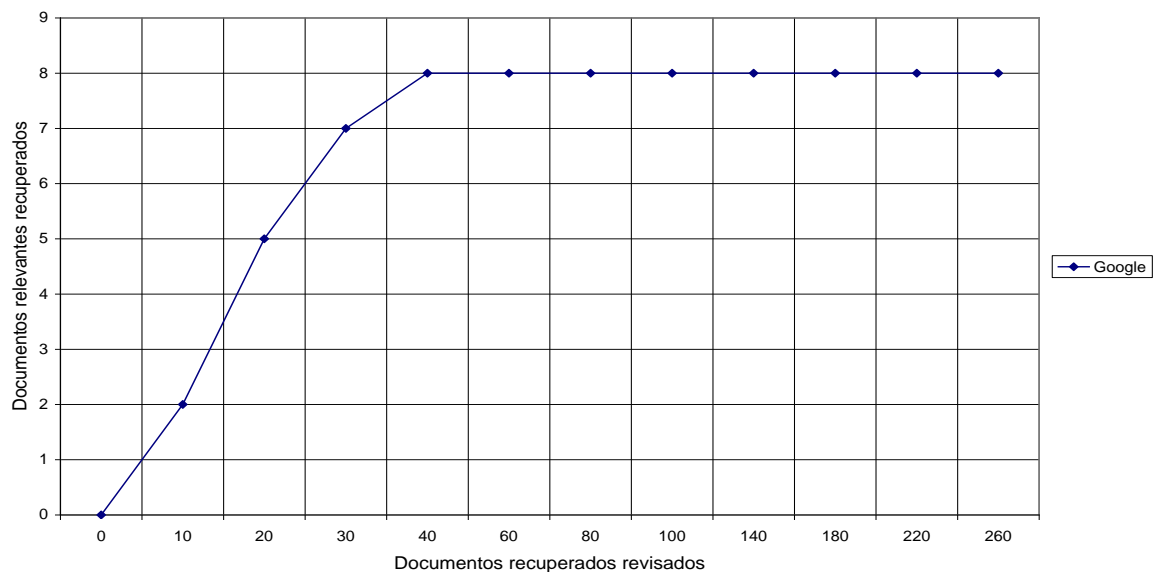


Figura 48. Resultados prueba 1 para la categoría mortero. Documentos relevantes recuperados vs. Documentos analizados manualmente.

## Análisis de resultados

La realización de este tipo de pruebas, las cuales no involucran el sistema multiagente desarrollado, sirve como base para determinar parámetros importantes para un óptimo desempeño del sistema multiagente.

El resultado arrojado tras la ejecución de esta prueba, muestra que en el 46% de las consultas realizadas sobre el motor de búsqueda google, más allá de 40 documentos recuperados no se hallaron resultados relevantes para el dominio. En el 23% de las consultas realizadas sobre el motor de búsqueda google, más allá de 60 documentos recuperados no se hallaron resultados relevantes en el 15% de las consultas realizadas sobre el motor más allá de 30 documentos recuperados no se hallaron resultados relevantes y en porcentajes del 7% más allá de 20 y 17 documentos recuperados no se hallaron documentos relevantes, en este último caso el motor no recuperó más que 17 documentos, lo que permite, para el caso particular del MAS, fijar un número máximo de 50 documentos a procesar por el MAS durante sus ejecuciones. Uno de los inconvenientes de generalizar este valor es en el caso en que mucho más allá de 260 documentos, aunque con posibilidades bajas, pero reales, existan más documentos relevantes para el dominio, en este caso el MAS estaría perdiendo la oportunidad de ampliar su índice y por consiguiente de entregar resultados más amplios al usuario.

## 6.2 PRUEBA 2

Titulo de la prueba: Capacidad de filtrado del sistema

### Objetivo:

Determinar la capacidad del sistema para la categorización automática de documentos en el dominio Mampostería Estructural, tomando como base la categorización manual de documentos realizada previamente bajo la supervisión del experto en el dominio.

### Estrategia de ejecución:

Para cada una de las trece categorías temáticas en que se divide el dominio, se selecciona un grupo de consultas a ser efectuadas sobre el motor de búsqueda, de manera que los documentos recuperados de esta consulta sean analizados, validados y categorizados de forma automática por el sistema multiagente.

Los resultados de este proceso se comparan con la categorización manual realizada sobre los resultados del mismo grupo de consultas al motor de búsqueda, tomando como criterio de comparación la selección previa de documentos realizada bajo la supervisión del experto en el dominio.

Escenario:

Datos de entrada: consultas previamente definidas para cada categoría.

Motor de búsqueda utilizado para la prueba: Google.

Número de categorías evaluadas: 13

Número de consultas: variable de acuerdo con la categoría analizada

Número de páginas analizadas por consulta: 50

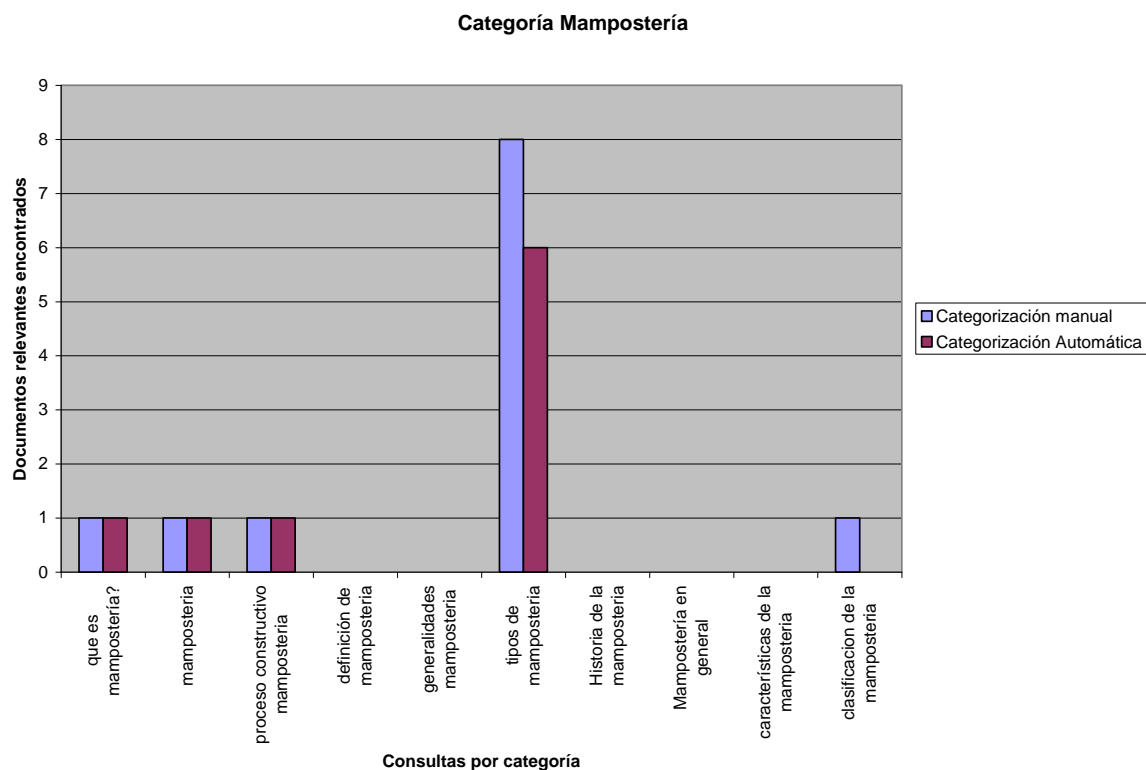
Se analizan 50 documentos por consulta, teniendo en cuenta los resultados de la prueba de efectividad del motor de búsqueda, en la cual se encontró que más allá de los 50 primeros resultados devueltos, el número de documentos relevantes recuperados en el dominio tiende a ser constante.

Resultados Obtenidos:

Caso 1: Categoría Acabados

No	Consulta	Documentos analizados	Documentos relevantes encontrados en forma manual	Documentos relevantes categorizados automáticamente por el sistema
1	que es mampostería?	50	1	1
2	Mampostería	50	1	1
3	proceso constructivo mampostería	50	1	1
4	definición de mamposteria	50	0	0
5	Generalidades mamposteria	50	0	0
6	tipos de mamposteria	50	8	6
7	Historia de la mamposteria	50	0	0
8	Mampostería en general	50	0	0
9	Características de la mampostería	50	0	0
10	Clasificación de la mampostería	50	1	0

**Tabla 25** Documentos relevantes para la categoría Mampostería clasificados manualmente y automáticamente dada una serie de consultas.



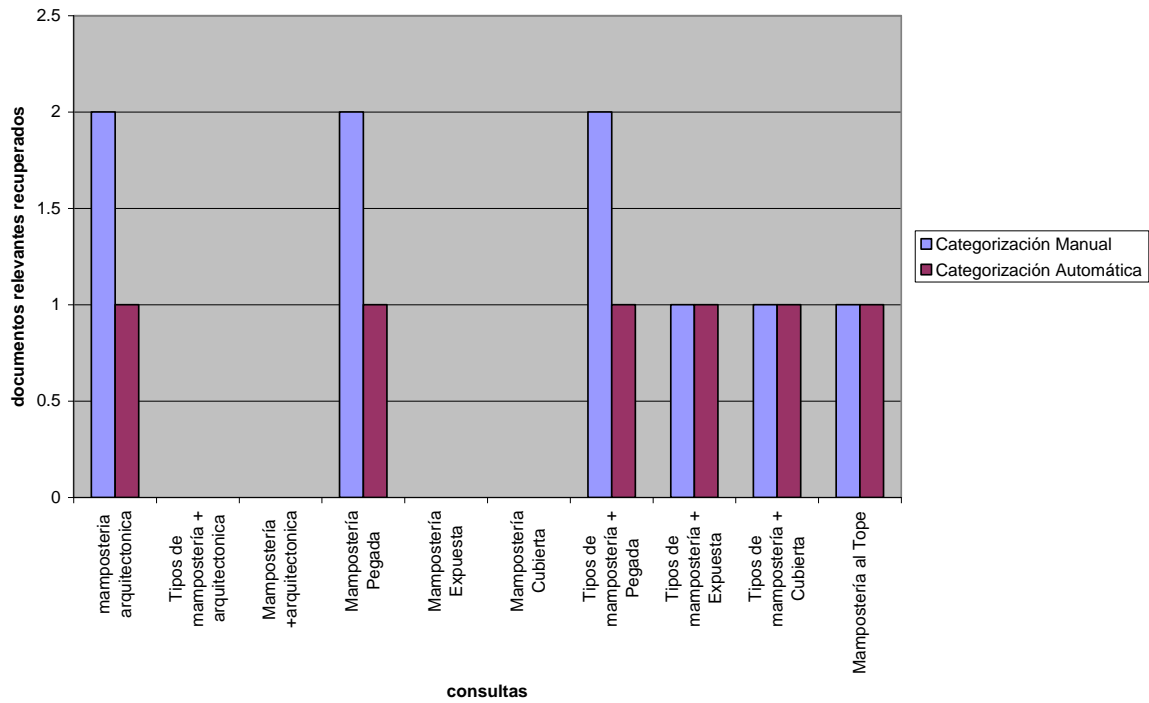
**Figura 49.** Documentos relevantes para la categoría Mampostería clasificados manualmente y automáticamente dada una serie de consultas

#### Caso 2: Categoría Mampostería Arquitectónica

No	Consulta	Documen- tos analizados	Documentos relevantes encontrados manualmente	Documentos relevantes filtrados automáticamente por el sistema
1	mampostería arquitectonica	50	2	1
2	Tipos de mampostería + arquitectonica	50	0	0
3	Mampostería +arquitectonica	50	0	0
4	Mampostería Pegada	50	2	1
5	Mampostería Expuesta	50	0	0
6	Mampostería Cubierta	50	0	0
7	Tipos de mampostería + Pegada	50	2	1
8	Tipos de mampostería + Expuesta	50	1	1
9	Tipos de mampostería + Cubierta	50	1	1
10	Mampostería al Tope	50	1	1

**Tabla 26.** Documentos relevantes para la categoría Mampostería arquitectónica clasificados manualmente y automáticamente dada una consulta

**Categoría: Mampostería Arquitectónica**



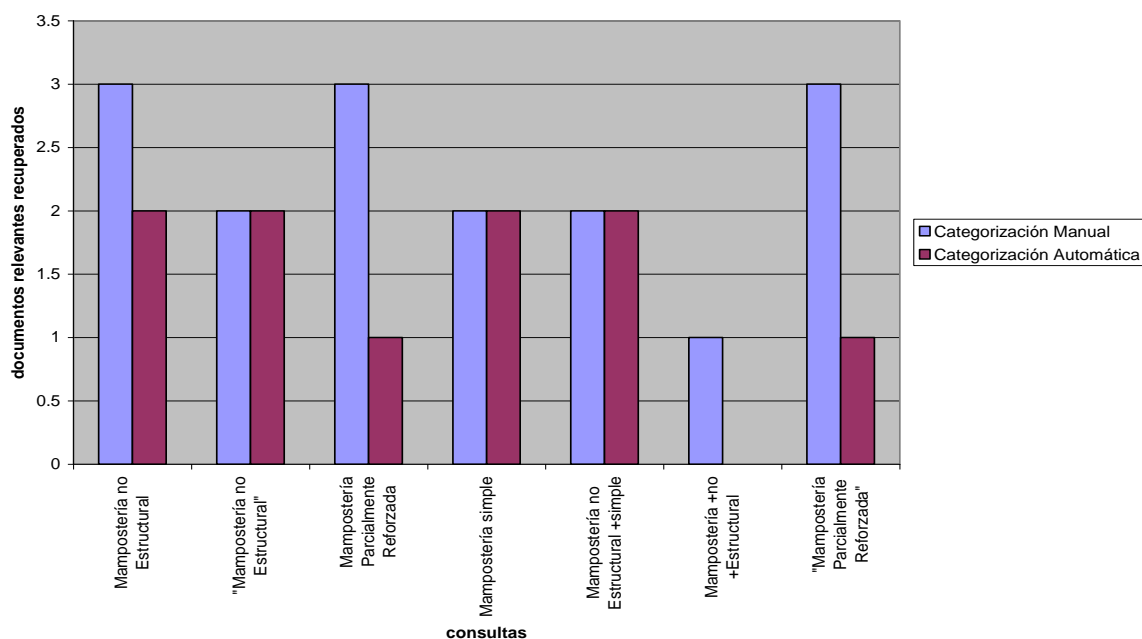
**Figura 50.** Documentos relevantes para la categoría Mampostería Arquitectónica clasificados manualmente y automáticamente dada una serie de consultas

**Caso 3: Categoría Mampostería no Estructural**

No	Consulta	Documen- tos analizados	Documentos relevantes encontrados manualmente	Documentos relevantes filtrados automáticamente por el sistema
1	Mampostería no Estructural	50	3	2
2	"Mampostería no Estructural"	50	2	2
3	Mampostería Parcialmente Reforzada	50	3	1
4	Mampostería simple	50	2	2
5	Mampostería no Estructural +simple	50	2	2
6	Mampostería +no +Estructural	50	1	0
7	"Mampostería Parcialmente Reforzada"	50	3	1

**Tabla 27.** Documentos relevantes para la categoría Mampostería no estructural clasificados manualmente y automáticamente dada una consulta

**Categoría: Mampostería no Estructural**



**Figura 51.** Documentos relevantes para la categoría Mampostería no estructural clasificados manualmente y automáticamente dada una serie de consultas

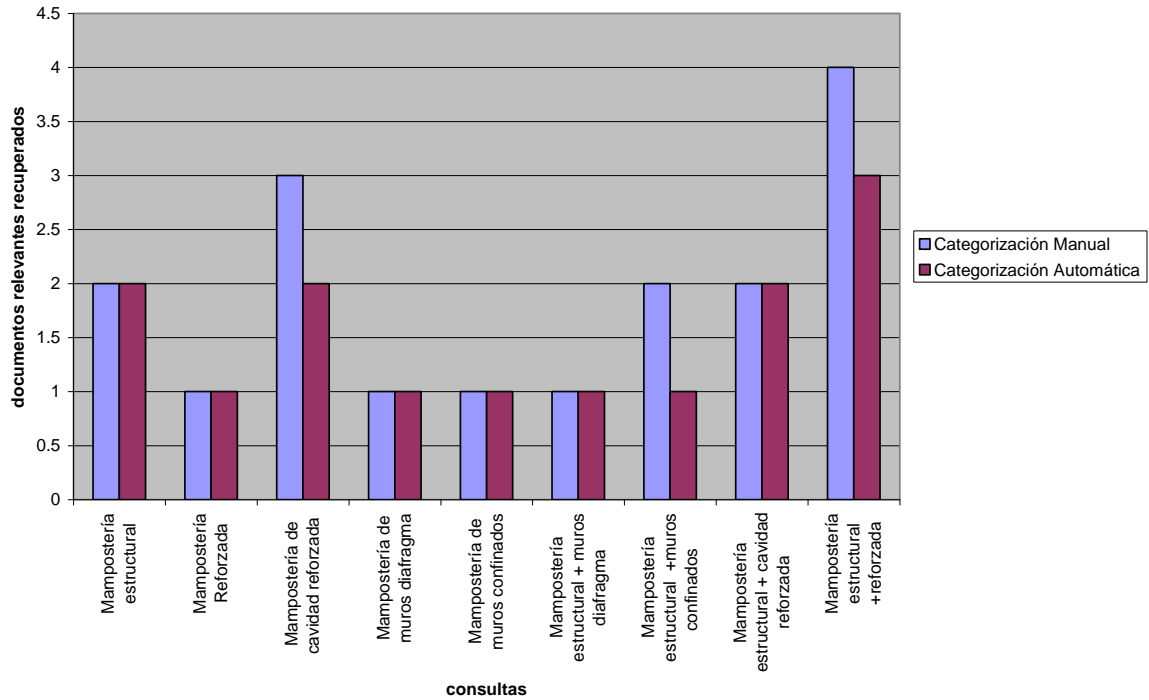
**Caso 4: Categoría Mampostería Estructural**

No	Consulta	Documentos analizados	Documentos relevantes encontrados manualmente	Documentos relevantes filtrados automáticamente por el sistema
1	Mampostería estructural	50	2	2
2	Mampostería Reforzada	50	1	1
3	Mampostería de cavidad reforzada	50	3	2
4	Mampostería de muros diafragma	50	1	1
5	Mampostería de muros confinados	50	1	1
6	Mampostería estructural + muros diafragma	50	1	1
7	Mampostería estructural +muros confinados	50	2	1
8	Mampostería estructural + cavidad reforzada	50	2	2
9	Mampostería estructural +reforzada	50	4	3

**Tabla 28.** Documentos relevantes para la categoría Mampostería estructural clasificados manualmente y automáticamente dada una consulta



**Categoría: Mampostería Estructural**



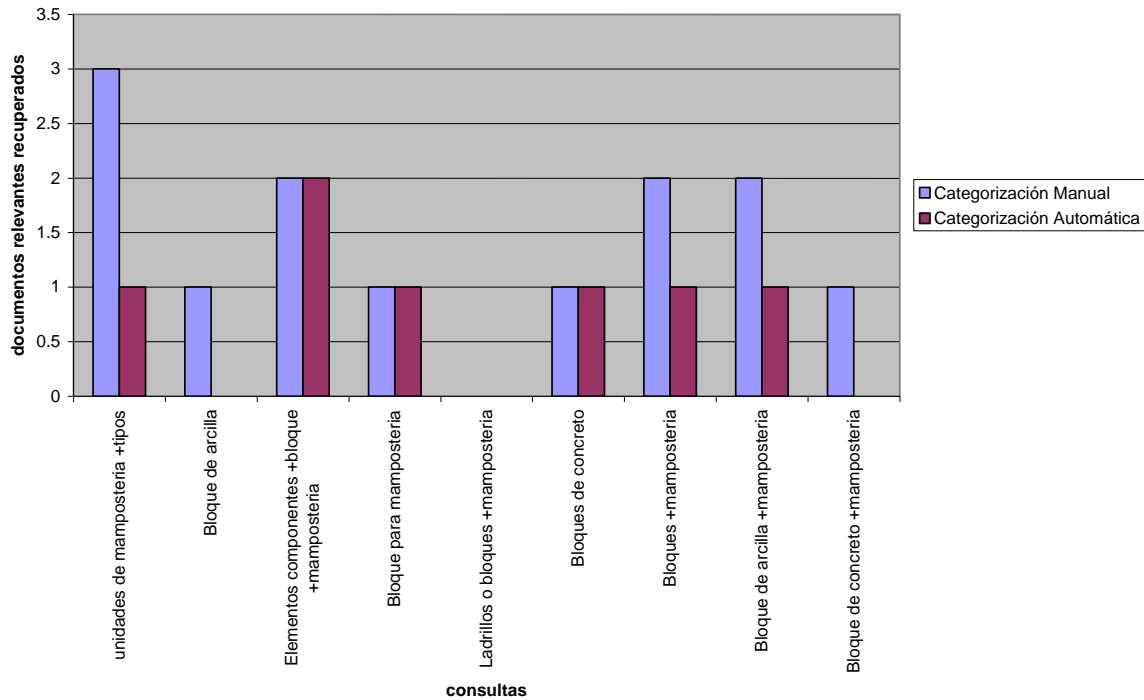
**Figura 52.** Documentos relevantes para la categoría Mampostería estructural clasificados manualmente y automáticamente dada una serie de consultas

**Caso 5: Categoría Unidad de Mampostería**

No	Consulta	Documentos analizados	Documentos relevantes encontrados manualmente	Documentos relevantes filtrados automáticamente por el sistema
1	unidades de mamposteria +tipos	50	3	1
2	Bloque de arcilla	50	1	0
3	Elementos componentes +bloque +mamposteria	50	2	2
4	Bloque para mamposteria	50	1	1
5	Ladrillos o bloques +mamposteria	50	0	0
6	Bloques de concreto	50	1	1
7	Bloques +mamposteria	50	2	1
8	Bloque de arcilla +mamposteria	50	2	1
9	Bloque de concreto +mamposteria	50	1	0

**Tabla 29.** Documentos relevantes para la categoría Unidad de Mampostería clasificados manualmente y automáticamente dada una consulta

**Categoría: Unidad de Mampostería**

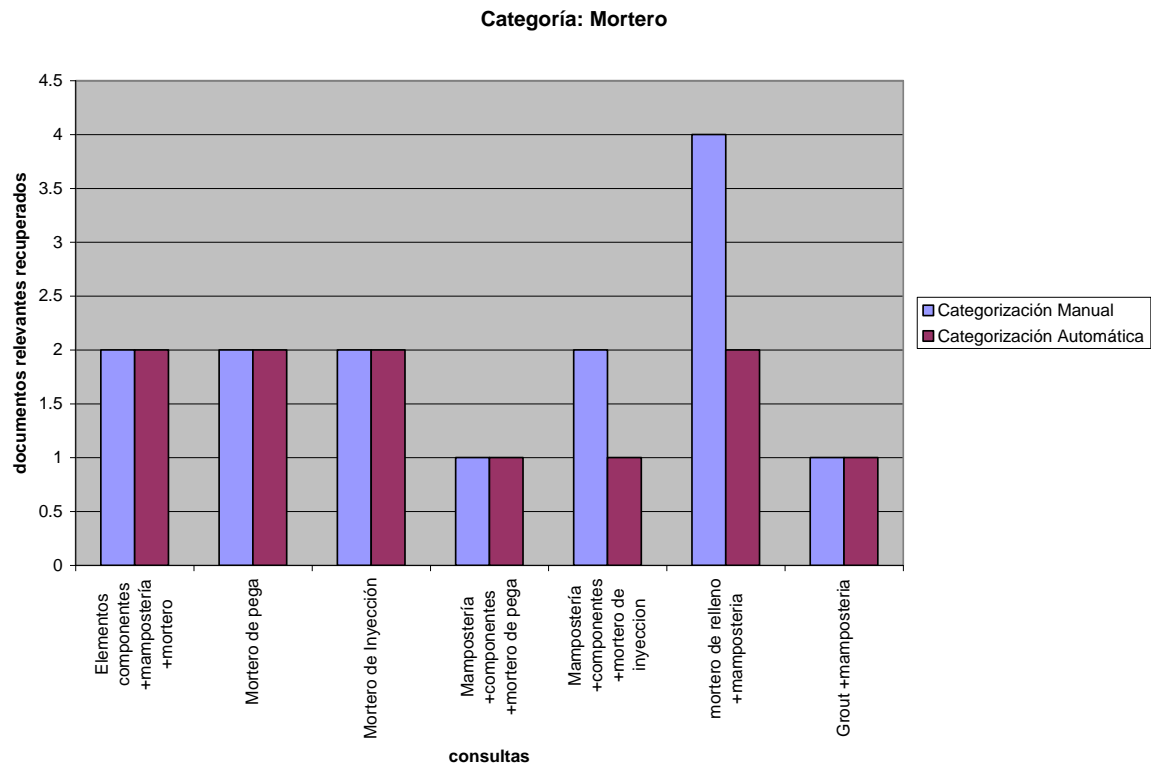


**Figura 53.** Documentos relevantes para la categoría Unidad de Mampostería clasificados manualmente y automáticamente dada una serie de consultas

**Caso 6: Categoría Mortero**

No	Consulta	Documen- tos analizados	Documentos relevantes encontrados manualmente	Documentos relevantes filtrados automáticamente por el sistema
1	Elementos componentes +mampostería +mortero	50	2	2
2	Mortero de pega	50	2	2
3	Mortero de Inyección	50	2	2
4	Mampostería +componentes +mortero de pega	50	1	1
5	Mampostería +componentes +mortero de inyeccion	50	2	1
6	mortero de relleno +mampostería	50	4	2
7	Grout +mampostería	50	1	1

**Tabla 30.** Documentos relevantes para la categoría Mortero clasificados manualmente y automáticamente dada una consulta

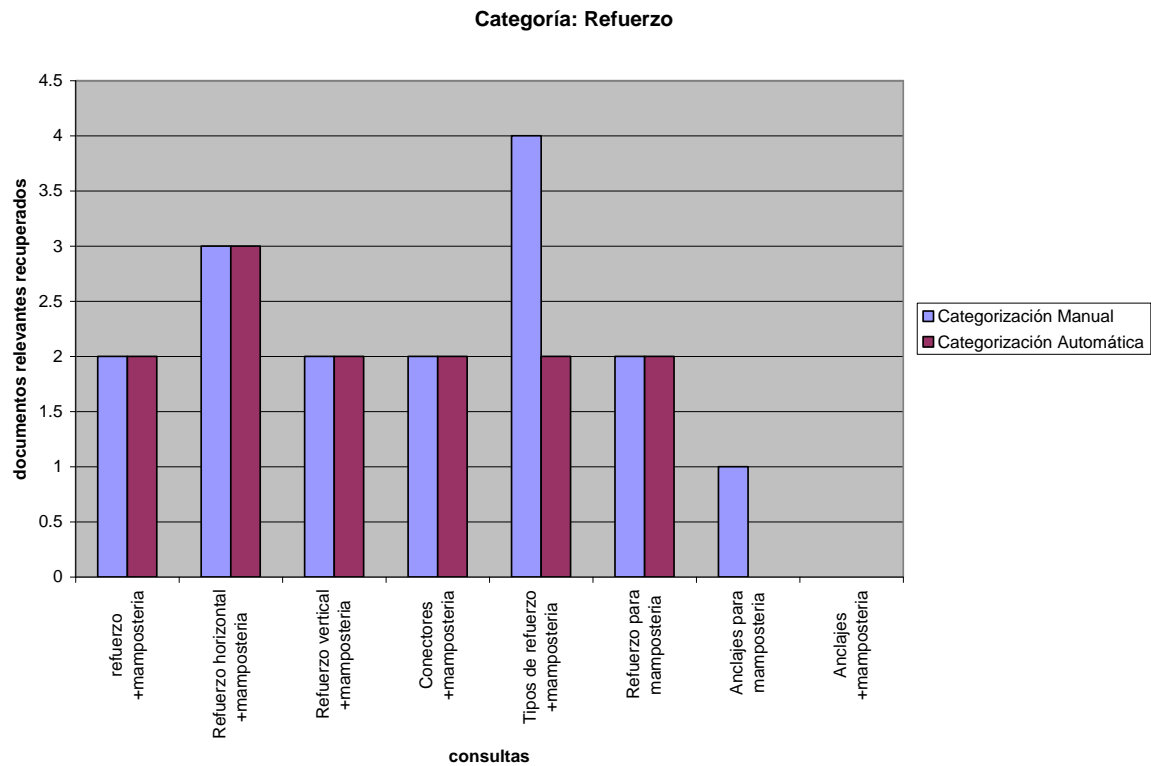


**Figura 54.** Documentos relevantes para la categoría Mortero clasificados manualmente y automáticamente dada una serie de consultas

#### Caso 7: Categoría Refuerzo

No	Consulta	Documen- tos analizados	Documentos relevantes encontrados manualmente	Documentos relevantes filtrados automáticamente por el sistema
1	refuerzo +mamposteria	50	2	2
2	Refuerzo horizontal +mamposteria	50	3	3
3	Refuerzo vertical +mamposteria	50	2	2
4	Conectores +mamposteria	50	2	2
5	Tipos de refuerzo +mamposteria	50	4	2
6	Refuerzo para mamposteria	50	2	2
7	Anclajes para mamposteria	50	1	0
8	Anclajes +mamposteria	50	0	0

**Tabla 31.** Documentos relevantes para la categoría Refuerzo clasificados manualmente y automáticamente dada una consulta

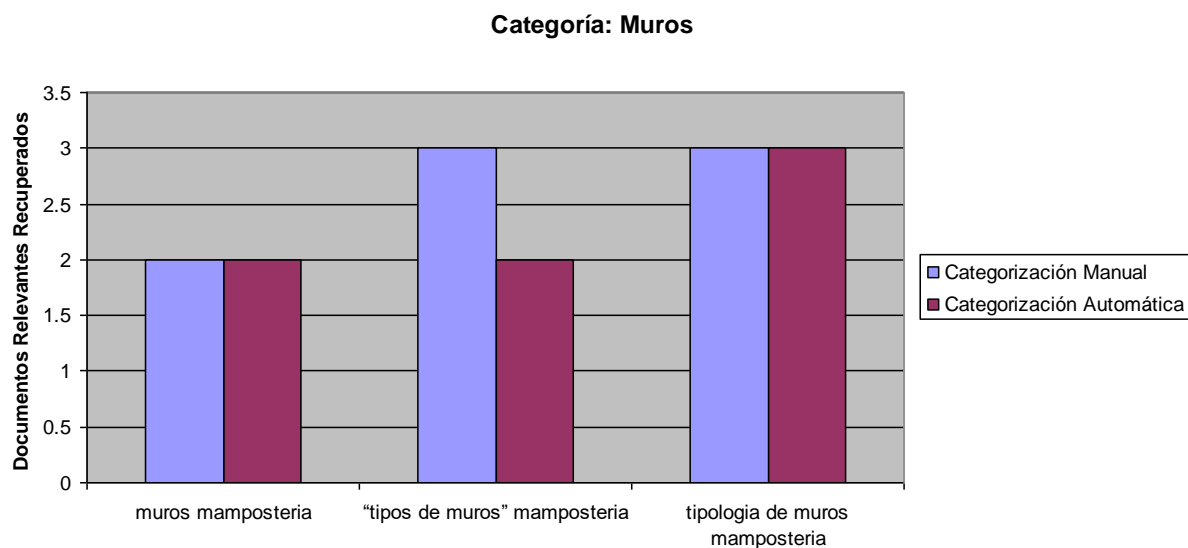


**Figura 55.** Documentos relevantes para la categoría Refuerzo clasificados manualmente y automáticamente dada una serie de consultas.

#### Caso 8: Categoría Muro

No	Consulta	Documentos analizados	Documentos relevantes encontrados manualmente	Documentos relevantes filtrados automáticamente por el sistema
1	muros mamposteria	50	2	2
2	“tipos de muros” mamposteria	50	3	2
3	tipologia de muros mamposteria	35	3	3

**Tabla 32.** Documentos relevantes para la categoría Muro clasificados manualmente y automáticamente dada una consulta

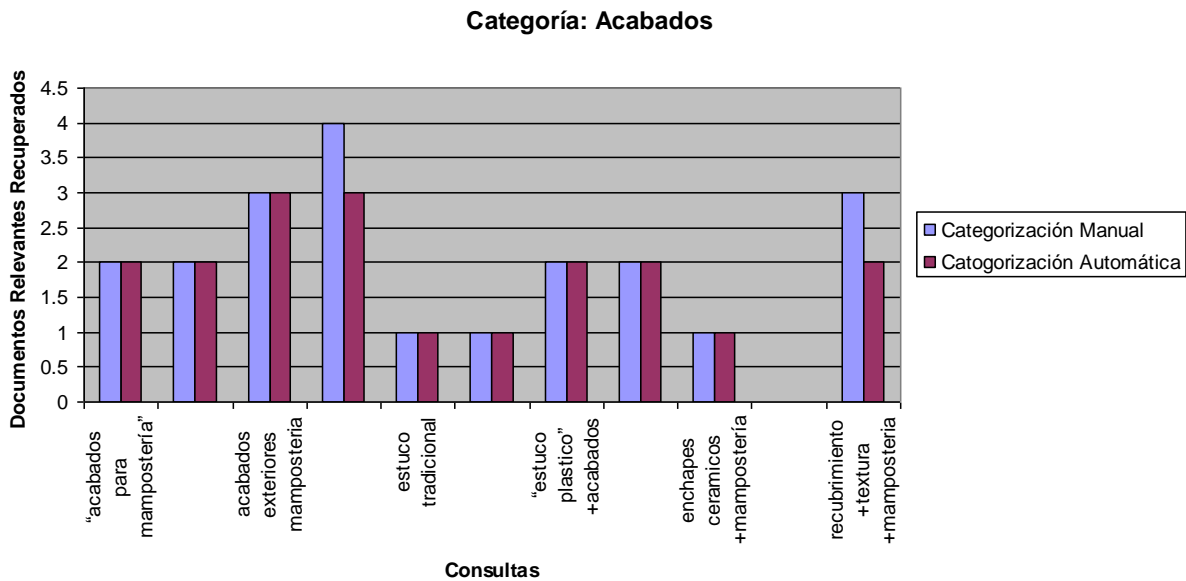


**Figura 56.** Documentos relevantes para la categoría Muro clasificados manualmente y automáticamente dada una serie de consultas.

#### Caso 9: Categoría Acabados

No	Consulta	Documentos analizados	Documentos relevantes encontrados manualmente	Documentos relevantes filtrados automáticamente por el sistema
1	"acabados para mampostería"	50	2	2
2	acabados interiores mamposteria	50	2	2
3	acabados exteriores mamposteria	50	3	3
4	revoque + "acabados para interiores"	50	4	3
5	estuco tradicional	50	1	1
6	estuco clasico	50	1	1
7	"estuco plastico" +acabados	50	2	2
8	"tipos de pintura" mamposteria	50	2	2
9	enchapes ceramicos +mampostería +acabados	50	1	1
10	"enchapes no ceramicos" +mampostería +acabados	0	0	0
11	recubrimiento +textura +mamposteria	50	3	2

**Tabla 33.** Documentos relevantes para la categoría Acabados clasificados manualmente y automáticamente dada una consulta



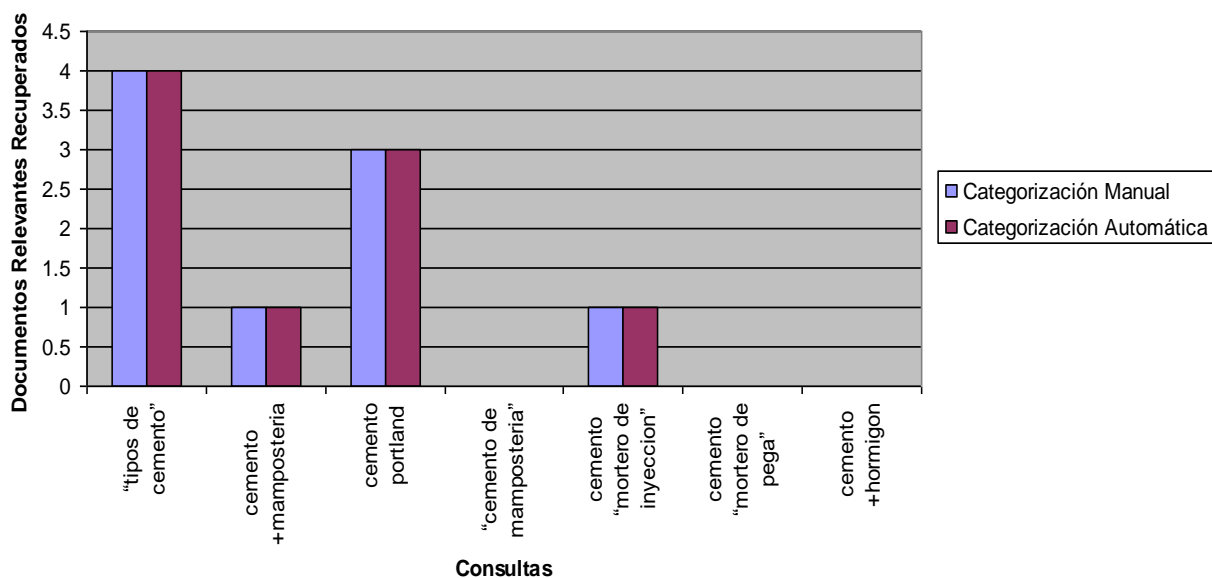
**Figura 57.** Documentos relevantes para la categoría Acabados clasificados manualmente y automáticamente dada una serie de consultas.

#### Caso 10: Categoría Cemento

No	Consulta	Documentos analizados	Documentos relevantes encontrados manualmente	Documentos relevantes filtrados automáticamente por el sistema
1	"tipos de cemento"	50	4	4
2	cemento +mampostería	50	1	1
3	cemento portland	50	3	3
4	"cemento de mampostería"	39	0	0
5	cemento "mortero de inyeccion"	50	1	1
6	cemento "mortero de pega"	50	0	0
7	cemento +hormigón	50	0	0

**Tabla 34.** Documentos relevantes para la categoría Cemento clasificados manualmente y automáticamente dada una consulta

### Categoría: Cemento

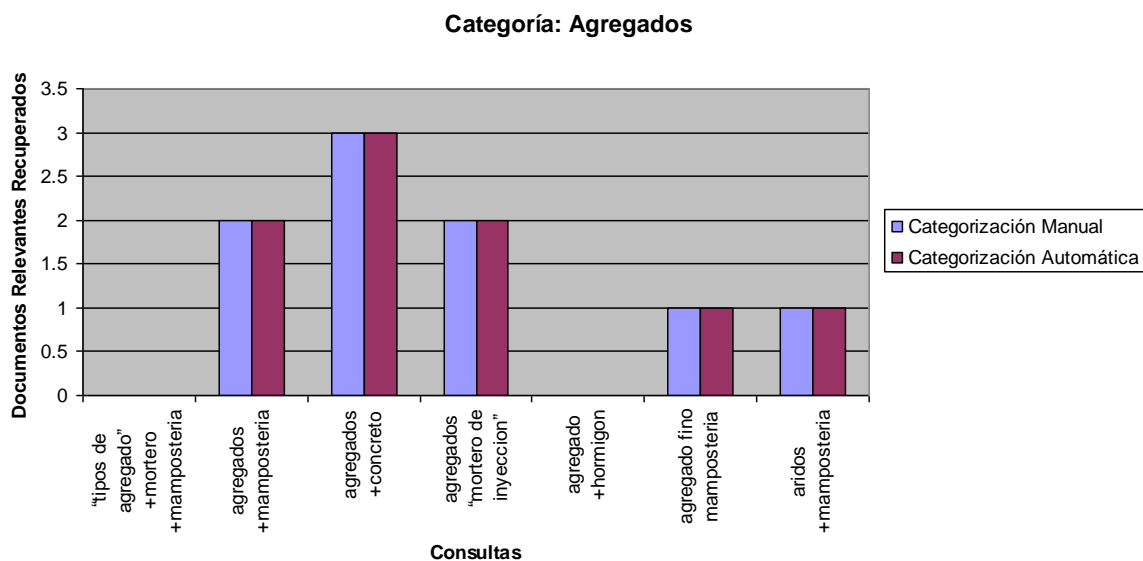


**Figura 58.** Documentos relevantes para la categoría Cemento clasificados manualmente y automáticamente dada una serie de consultas.

### Caso 11: Categoría: Agregados

No	Consulta	Documentos analizados	Documentos relevantes encontrados manualmente	Documentos relevantes filtrados automáticamente por el sistema
1	"tipos de agregado" +mortero +mamposteria	50	0	0
2	agregados +mamposteria	50	2	2
3	agregados +concreto	50	3	3
4	agregados "mortero de inyeccion"	31	2	2
5	agregado +hormigon	50	0	0
6	agregado fino mamposteria	50	1	1
7	aridos +mamposteria	50	1	1

**Tabla 35.** Documentos relevantes para la categoría Agregados clasificados manualmente y automáticamente dada una consulta



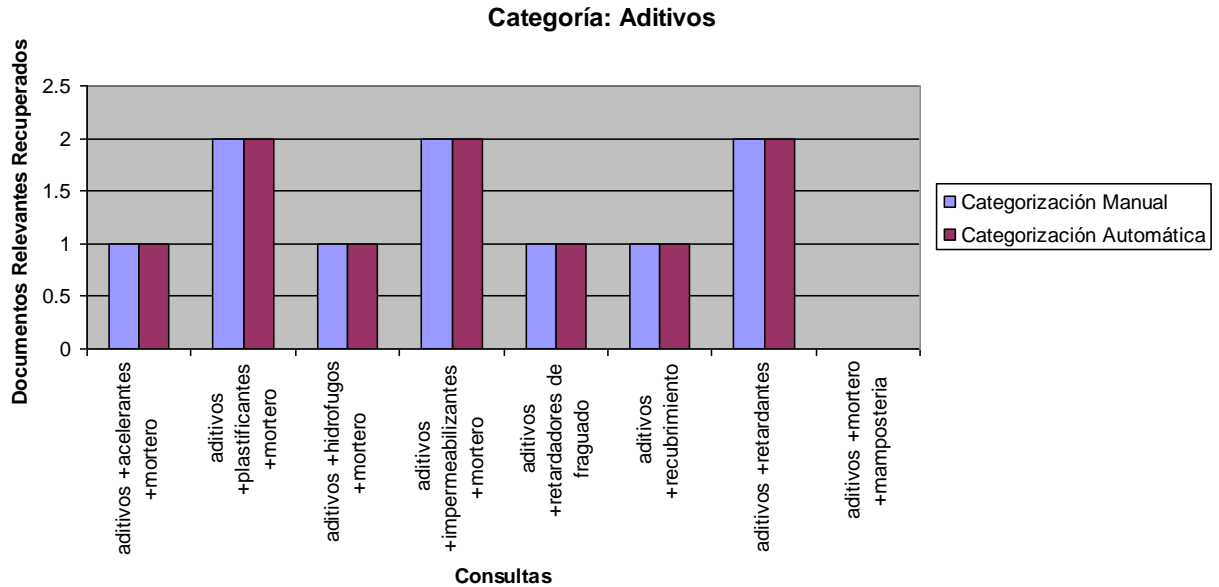
**Figura 59.** Documentos relevantes para la categoría Agregados clasificados manualmente y automáticamente dada una serie de consultas.

#### Caso 12: Categoría Aditivos

No	Consulta	Documentos analizados	Documentos relevantes encontrados manualmente	Documentos relevantes filtrados automáticamente por el sistema
1	aditivos +acelerantes +mortero	50	1	1
2	aditivos +plastificantes +mortero	50	2	2
3	aditivos +hidrofugos +mortero	35	1	1
4	aditivos +impermeabilizantes +mortero	50	2	2
5	aditivos +retardadores de fraguado	50	1	1
6	aditivos +recubrimiento	50	1	1
7	aditivos +retardantes	50	2	2
8	aditivos +mortero +mamposteria	50	0	0

**Tabla 36.** Documentos relevantes para la categoría Aditivos clasificados manualmente y automáticamente dada una consulta



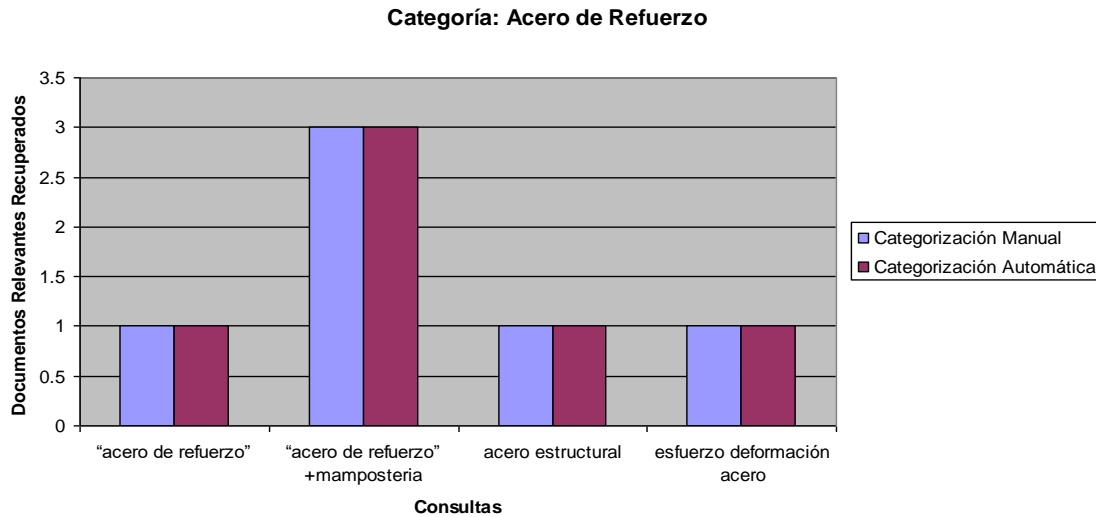


**Figura 60.** Documentos relevantes para la categoría Aditivos clasificados manualmente y automáticamente dada una serie de consultas.

#### Caso 13: Categoría Acero de Refuerzo

No	Consulta	Documentos analizados	Documentos relevantes encontrados manualmente	Documentos relevantes filtrados automáticamente por el sistema
1	“acero de refuerzo”	50	1	1
2	“acero de refuerzo” + mamposteria	50	3	3
3	Acero estructural	35	1	1
4	esfuerzo deformación acero	50	1	1

**Tabla 37.** Documentos relevantes para la categoría Acero de refuerzo clasificados manualmente y automáticamente dada una consulta



**Figura 61.** Documentos relevantes para la categoría Acero de refuerzo clasificados manualmente y automáticamente dada una serie de consultas.

#### Análisis de resultados:

El resultado de este grupo de pruebas permite demostrar que la capacidad de filtrado del sistema desarrollado es alta ya que al compararlo con el filtrado manual realizado previamente resultan muy similares.

El resultado arrojado tras la ejecución de esta prueba, muestra que el 79.8% de los documentos categorizados de forma automática por el sistema, fueron ubicados en la misma categoría que la seleccionada en el proceso manual de categorización.

En el 20.2% restante, el margen de error en la categorización no supera el 25% excepto en los casos en que el número de documentos a categorizar es muy bajo (menor a 2 documentos).

Los resultados obtenidos confirman la superioridad del proceso de categorización realizado por seres humanos, sin embargo la diferencia con la categorización automática realizada por el sistema es pequeña y en algunas categorías llega a ser igual.

### 6.3 PRUEBA 3

Título de la Prueba: Precisión porcentual de Google vs. Sistema de Optimización de Consultas

Objetivo:

Realizar un análisis comparativo entre la precisión de los resultados arrojados por google frente a la precisión de los resultados arrojados por el sistema, medidos en términos de documentos relevantes en el dominio frente a documentos no relevantes.

Estrategia de ejecución:

Plantear una consulta de usuario relevante en el dominio mampostería estructural.

Ejecutar la consulta sobre el motor de búsqueda google y sobre el MAS.

Determinar el número máximo de documentos que se van a analizar.

El número de documentos a analizar debe ser igual tanto para google como para el MAS.

Se determina en bloques de documentos recuperados por los dos sistemas, cuantos documentos relevantes en el dominio se recuperaron y se calcula la proporción entre documentos relevantes recuperados y documentos recuperados individualmente por cada uno de los dos sistemas

Escenario:

Datos de entrada: consultas previamente definidas para cada categoría.

Sistemas utilizados para la prueba: Google y MAS

Número de categorías evaluadas: 13

Número de consultas: Quince consultas

Número de documentos a analizar: 30

Caso 1: Categoría Acabados

Consulta de usuario: acabados para mampostería

Número de documentos a analizar: 30

Los resultados obtenidos se resumen en la tabla número 38.

Google			MAS		
Primeros Documentos recuperados	Documentos relevantes	Precisión (%)	Primeros Documentos recuperados	Documentos relevantes	Precisión
10	2	20	10	8	80
20	3	15	20	8	40
30	4	13.33333333	30	8	26.6666667

Tabla 38. Resumen resultados prueba 3 para la categoría acabados.

Caso 2: Categoría Acero de refuerzo

Consulta de usuario: acero de refuerzo

Los resultados obtenidos se resumen en la tabla número 39.

Google			MAS		
Primeros Documentos recuperados	Documentos relevantes	Precisión (%)	Primeros Documentos recuperados	Documentos relevantes	Precisión
10	1	10	10	0	0
20	3	15	20	0	0
30	5	16.6666667	30	5	16.6666667

Tabla 39. Resumen resultados prueba 3 para la categoría acero de refuerzo

Caso 3: Categoría Aditivos

Consulta de usuario: aditivos mortero

Los resultados obtenidos se resumen en la tabla número 40.

Google			MAS		
Primeros Documentos recuperados	Documentos relevantes	Precisión (%)	Primeros Documentos recuperados	Documentos relevantes	Precisión
10	3	30	10	5	50
20	4	20	20	7	35
30	6	20	30	7	23.3333333

Tabla 40. Resumen resultados prueba 3 para la categoría aditivos.

Caso 4: Categoría Agregado

Consulta de usuario: tipos de agregados mortero mampostería

Los resultados obtenidos se resumen en la tabla número 41.

Google			MAS		
Primeros Documentos recuperados	Documentos relevantes	Precisión (%)	Primeros Documentos recuperados	Documentos relevantes	Precisión
10	4	40	10	2	20
20	5	25	20	5	25
30	5	16.6666667	30	5	16.6666667

Tabla 41. Resumen resultados prueba 3 para la categoría Agregado.

Caso 5: Categoría Cementos

Consulta de usuario: Cementos Mampostería

Los resultados obtenidos se resumen en la tabla número 42.

Google			MAS		
Primeros Documentos recuperados	Documentos relevantes	Precisión (%)	Primeros Documentos recuperados	Documentos relevantes	Precisión
10	4	40	10	0	0
20	7	35	20	7	35
30	8	26.6666667	30	7	23.3333333

Tabla 42. Resumen resultados prueba 3 para la categoría Cementos.

Caso 6: Categoría Mampostería

Consulta de usuario: Tipos de Mampostería

Los resultados obtenidos se resumen en la tabla número 43.

MAS			Google		
Primeros Documentos recuperados	Documentos relevantes	Precisión (%)	Primeros Documentos recuperados	Documentos relevantes	Precisión
10	3	30	10	0	0
20	7	35	20	7	35
30	8	26.6666667	30	7	23.3333333

Tabla 43. Resumen resultados prueba 3 para la categoría Mampostería.

Caso 7: Categoría Mampostería Arquitectónica

Consulta de usuario: Mampostería Arquitectónica

Los resultados obtenidos se resumen en la tabla número 44.

Google			MAS		
Primeros Documentos recuperados	Documentos relevantes	Precisión (%)	Primeros Documentos recuperados	Documentos relevantes	Precisión
10	2	20	10	0	0
20	3	15	20	2	10
30	-	-	30	2	6.66666667

Tabla 44. Resumen resultados prueba 3 para la categoría mampostería arquitectónica.

Caso 8: Mampostería Estructural

Consulta de usuario: Mampostería Estructural

Los resultados obtenidos se resumen en la tabla número 45.

Google			MAS		
Primeros Documentos recuperados	Documentos relevantes	Precisión (%)	Primeros Documentos recuperados	Documentos relevantes	Precisión
10	2	20	10	0	0
20	3	15	20	0	0
30	4	13.33333333	30	6	20

Tabla 45. Resumen resultados prueba 3 para la categoría mampostería estructural

Caso 9: Mampostería No Estructural

Consulta de usuario: acabados para Mampostería No Estructural

Los resultados obtenidos se resumen en la tabla número 46

Google			MAS		
Primeros Documentos recuperados	Documentos relevantes	Precisión (%)	Primeros Documentos recuperados	Documentos relevantes	Precisión
10	2	20	10	0	0
20	4	20	20	0	0
30	5	16.66666667	30	6	20

Tabla 46. Resumen resultados prueba 3 para la categoría mampostería no estructural

Caso 10: Categoría Mortero

Consulta de usuario: mortero mampostería

Los resultados obtenidos se resumen en la tabla número 47.

Google			MAS		
Primeros Documentos recuperados	Documentos relevantes	Precisión (%)	Primeros Documentos recuperados	Documentos relevantes	Precisión
10	3	30	10	0	0
20	4	20	20	1	5
30	6	20	30	9	30

Tabla 47. Resumen resultados prueba 3 para la categoría mortero.

Caso 11: Categoría Muro

Consulta de usuario: tipologías de muros mampostería

Los resultados obtenidos se resumen en la tabla número 48

Google			MAS		
Primeros Documentos recuperados	Documentos relevantes	Precisión (%)	Primeros Documentos recuperados	Documentos relevantes	Precisión
10	1	10	10	6	60
20	3	15	20	6	30
30	5	16.6666667	30	6	20

Tabla 48. Resumen resultados prueba 3 para la categoría Muro.

Caso 12: Categoría Refuerzo

Consulta de usuario: tipos de refuerzo mampostería

Los resultados obtenidos se resumen en la tabla número 49.

Google			MAS		
Primeros Documentos recuperados	Documentos relevantes	Precisión (%)	Primeros Documentos recuperados	Documentos relevantes	Precisión
10	2	20	10	4	40
20	3	15	20	5	25
30	4	13.3333333	30	5	16.6666667

Tabla 49. Resumen resultados prueba 3 para la categoría Refuerzo.

Caso 13: Categoría Unidad Mampostería

Consulta de usuario: tipos de refuerzo mampostería

Los resultados obtenidos se resumen en la tabla número 50.

Google			MAS		
Primeros Documentos recuperados	Documentos relevantes	Precisión (%)	Primeros Documentos recuperados	Documentos relevantes	Precisión
10	2	20	10	8	80
20	5	25	20	8	40
30	7	23.3333333	30	8	26.6666667

Tabla 50.. Resumen resultados prueba 3 para la categoría Unidad Mampostería.

Caso 14: Categoría Consulta típica de usuario: Pinturas para exteriores en muros de mampostería.

Los resultados obtenidos se resumen en la tabla número 51.

Google			MAS		
Primeros Documentos recuperados	Documentos relevantes	Precisión (%)	Primeros Documentos recuperados	Documentos relevantes	Precisión
10	2		10	8	
20	2		20	8	
30	2		30	8	

Tabla 52. Resumen resultados prueba 3 para la categoría acabados.

Caso 15: Categoría: Consulta que no corresponde al dominio

Consulta de usuario: Edificios apoticados en Colombia

Los resultados obtenidos se resumen en la tabla número 53.

Google			MAS		
Primeros Documentos recuperados	Documentos relevantes	Precisión (%)	Primeros Documentos recuperados	Documentos relevantes	Precisión
10	1		10	0	
20	0		20	0	
30	0		30	0	

Tabla 53. Resumen resultados prueba 3 para la categoría no valida.



## Análisis de Resultados

En las tablas 54., y 55., se resumen algunas variables estadísticas discriminados por el número de resultados analizados para los dos sistemas.

<b>Número de Resultados</b>	<b>10</b>	<b>20</b>	<b>30</b>
<b>Máximo %</b>	40	35	26.6666667
<b>Mínimo %</b>	10	15	13.3333333
<b>Promedio %</b>	23.8461538	20.7692308	18.6111111
<b>Desviación</b>	9.60768923	7.31612199	4.81125224

**Tabla 54.** Resumen de Resultados Google

<b>Número de Resultados</b>	<b>10</b>	<b>20</b>	<b>30</b>
<b>Máximo</b>	80	40	30
<b>Mínimo</b>	0	0	6.6666667
<b>Promedio</b>	25.3846154	21.5384615	20.7692308
<b>Desviación</b>	32.3046397	16.1225276	5.95567674

**Tabla 55.** Resumen de Resultados MAS

De la cuales se resaltan las siguientes observaciones:

El MAS obtuvo un mejor resultado promedio de la precisión porcentual en las consultas realizadas, lo cual es un indicativo de que los resultados obtenidos por el usuario son adecuados.

El MAS presentó valores de precisión máximos, más elevados de los que presentó Google, sin embargo Google presentó valores mínimos de precisión más elevados que el MAS, lo cual indica que el motor de búsqueda le garantiza al usuario un mínimo de precisión en sus resultados, sin embargo el MAS presenta mayor precisión en los casos en que entrega buenos resultados

Si bien los documentos analizados en promedio fueron de 30 en ambos sistemas, google ofrece una cantidad mucho mayor de referencias, esta aparente ventaja es opacada por los menores niveles de precisión presentados en comparación con MAS, aun siendo un sistema con una colección mucho más pequeña de documentos

Es importante resaltar que las consultas típicas de usuario, al estar compuesta de un número mayor de términos, facilitan el mejor desempeño del sistema en términos de precisión.

## 7 CONCLUSIONES

Uno de los problemas fundamentales de la web es su tamaño, heterogeneidad e inconsistencia; los recursos cambian y se multiplican y la velocidad de respuesta de los motores de búsqueda no está ligada, generalmente, a una efectiva recuperación de información” [OPP, 2000]. El sistema multiagente (MAS) desarrollado plantea una solución que, como prototipo, realiza una buena organización de la información en el dominio mampostería estructural, gracias al uso de tecnologías como las ontologías que permiten representar el conocimiento en el dominio de una manera estructurada.

El uso de modelos matemáticos como el modelo de espacio vectorial aplicado en los procesos de recuperación de información en la web, permiten realizar una buena representación de la información, sin embargo, dichos modelos no son suficientes para manejar de forma homogénea la gran variedad de formatos por medio de los cuales se presenta la información en la web

Técnicas adecuadas de modelado del conocimiento como las ontologías, nos acercan a una mejor representación de gran variedad de información, sin embargo, la poca utilización que se hace actualmente de los lenguajes y estándares disponibles para la representación de la información en la web, dificulta la explotación de las ontologías en procesos como el de recuperación de información relevante en la web

Las capacidades que ofrecen los agentes inteligentes en la ejecución de procesos tan complejos como el de recuperación de información relevante en la web se complementan muy bien con las ventajas de una adecuada representación de la información

Es de particular importancia notar que el sistema de optimización de búsquedas en el dominio mampostería estructural, trabaja como un sistema muy selectivo de información. Una vez el usuario entrega su consulta, el sistema ordena las categoría para las cuales esta consulta tiene mayor relación y posteriormente entrega, en el orden en que clasificó la consulta, bloques completos de los documento de la colección que se asocian a las

categoría. Con ello se consigue o una satisfacción plena de la consulta, o un bajo nivel de precisión en los resultados. Si bien es una propuesta riesgosa, la interfaz de usuario presenta otras alternativas para mitigar dicho riesgo, como son las búsquedas rápidas sobre las categorías.

El sistema de filtrado de información desarrollado y expuesto en este documento es un gran paso en el desarrollo de proyectos que abordan nuevas e interesantes temáticas y tecnologías como los sistemas multiagente y los sistemas de representación del conocimiento y se constituye en un aporte innovador que demuestra el gran potencial que poseen estas tecnologías.

El resultado de este trabajo demuestra que si es posible, en la web actual, mediante procesos automáticos, recuperar información relevante evitando al usuario invaluables pérdidas de tiempo en procesos de selección y evaluación de documentos.

El sistema de filtrado desarrollado es una herramienta muy útil para los usuarios interesados en la temática mampostería estructural, puesto que, recupera información relevante en la web, en donde, además de ser escasa, se encuentra dispersa y desordenada y la ordena por categorías, para ofrecer a estos usuarios información de calidad, ordenada y fácilmente accesible.

## **8. TRABAJO FUTURO**

El sistema de optimización de búsquedas en el dominio mampostería estructural integra una serie de tecnologías complejas, cada una de ellas con características idóneas para la solución de problemas específicos cuyo estudio a mayor profundidad permitirá enriquecer el sistema actual mediante la implementación de mejoras como las siguientes:

Implementar un modelo de representación genérico de las consultas de usuario que permita desligar el sistema de filtrado del dominio modelado, y conseguir así una aplicación genérica que pueda ser usada en cualquier dominio.

Implementar procesos de refinamiento automático de los esquemas de categorización de documentos que permitan ir refinando los vectores óptimos en la medida en que se recuperan nuevos documentos relevantes para las distintas categorías

Implementar métodos de aprendizaje de los agentes de tal forma que puedan mejorar su desempeño en la categorización de documentos y consultas a partir de las ejecuciones anteriores.

Mejorar el proceso de optimización realimentando el sistema con información proveniente de los usuarios, mediante el uso de técnicas como la realimentación por relevancia que permitan tener en cuenta aspectos como los intereses particulares de los usuarios y el grado de satisfacción del usuario con el sistema.

Integrar al sistema otros proveedores de información diferentes a google y de esta manera cubrir en el proceso de recuperación de información, un porcentaje más alto de los recursos que alberga actualmente la web.

## 9 REFERENCIAS

[AKW, 2005] Akwan Information Technologies: la compra brasileña de Google [En línea]. Brasil. 2005. <<http://www.euroresidentes.com/Blogs/internet/2005/07/akwan-la-compra-brasilea-de-google.html>, <http://www.todobr.com.br/>>

[ALC, 2004] Alcaide, R., García, A., Alcañiz, M. and Marcos, A. Sistemas de Representación y Procesamiento Automático del Conocimiento: Ontologías en la web semántica [En línea]. Valencia: Universidad Politécnica, Facultad de Informática. 2004 <[http://personales.upv.es/ccarrasc/doc/2003-2004/Onto\\_WS/OntologiasenlaWebSemantica.htm/](http://personales.upv.es/ccarrasc/doc/2003-2004/Onto_WS/OntologiasenlaWebSemantica.htm/)>

[ASF, 1999] ASF: Apache Software Foundation. The Apache Jakarta Project [En línea]. <<http://jakarta.apache.org/site/legal.html>>

[ASF, 2004] ASF: Apache Software Foundation. Apache License [En línea]. <<http://www.apache.org/licenses/>>

[BAE, 1999] Baeza-Yates, R. and Ribeiro-Neto, B.:Modern Information Retrieval. New York: ACM Press; Harlow [etc.] : Adisson-Wesley, 1999 XX, 513 p. ISBN 0-201-39829-X.

[BAE, 2000] Baeza-Yates, R. and Navarro, G. Barcino: Tecnología de búsquedas para empresas y particulares [En línea]. Chile: Universidad. <<http://www.akwan.cl/>>

[BAL, 1997] Balabanovic, M. and Fab, Y. Content-based, collabortive recommendation [En línea]. Puebla: UDLA-P. <[http://www.pue.udlap.mx/~tesis/lis/ramirez\\_v\\_m/](http://www.pue.udlap.mx/~tesis/lis/ramirez_v_m/)>

[BAL, 2004] Baloian, N., Galdames, P., Collazos, C. and Guerrero, L., "A Model for a Collaborative Recommender System for Multimedia Learning Material", Springer LNCS, Berlin Heidelberg, Germany. San Carlos: Accepted for 10th. International Workshop on Groupware (CRIWG'2004).

[BAR, 1999] Barry, L., Vinton G., Cerf, D., Clark, R., Kahn, L., Lynch, D., Postel, J., Lawrence G. and Roberts, S. Una breve historia de Internet (Primera Parte) [En línea]. <<http://www.ati.es/DOCS/internet/histint/histint1.html/>>

[BEN, 1979] Bennett, J. The commercial impact of usability in interactive systems: Management Computer Communication. Infotech State-of-the-Art. Infotech International. 1979.

[BER, 2001a] Berners-Lee, T. Director del consorcio World Wide Web. Investigador en el MIT y CSAL. Docente de ciencias de la computación Southampton ECS. <<http://www.w3.org/People/Berners-Lee/>>

[BER, 2001b] Berners-Lee, T., Hendler, J. and Lassila, O. The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities [En línea]. USA: Scientific American. < <http://www.scientificamerican.com/>>

[BERR, 2001] Berrocal, J., Figuerola, C., Zazo, A. and Rodríguez, E. Agentes Inteligentes: Recuperación Autónoma de Información en el Web [En línea]. Salamanca: Universidad, Departamento de Informática y Automática, Facultad de Documentación. 2001. <<http://reina.usal.es>>

[CAS, 2003a] Castells, P.: La Web Semántica [En línea]. Castilla: Universidad. 2003. <<http://www.ii.uam.es/~castells/publications/castells-uclm03.pdf>>

[CAS, 2003b] Pablo Castells, P. Aplicación de técnicas de la Web semántica [En línea]. Madrid: Universidad Autónoma, Escuela Politécnica Superior. <<http://giig.ugr.es/~mgea/coline02/Articulos/pcastells.pdf>>

[COL, 2004] Collazos, C., Vargas, M. and Vélez, A. TodoCo: Buscador Semántico Colombiano. Cauca: Universidad. Primeras Jornadas Web Semántica. 2004.

[DEC, 2002a] Decker, S., Lacher, M. and Staab, S. Proyecto OntoAgent [En línea]. Stanford: University. 2002. <<http://mondeca-publishing.com/s/anonymous/title10213.html>>

[DEC, 2002b] Decker, S. Proyecto Scalable Knowledge Composition (SKC) [En línea]. Stanford: University. 2002. <<http://mondeca-publishing.com/s/anonymous/title10211.html>>

[DIN, 2004] Dinos, J. Arquitectura de un sistema basado en agentes para la recuperación de metadatos RDF en base a una ontología de documentos. Puerto Rico: Universidad de Mayagüez Campus, Tesis de grado para optar al título de Maestro en Ciencias en Ingeniería de computadoras. 2004.

[ETA, 2004] Electronic Transactions on Intelligence Artificial [En línea]. Annual Journal Volumen. <<http://www.ida.liu.se/ext/etai/>>

[FER, 2002] Fernández, M. Historia de Internet, desde sus precursores hasta la actualidad [En línea]. <<http://www.southlink.com.ar/manual/www.htm/>>

[FIG, 2003] Figuerola, C., Berrocal, J., Zazo, A. and Rodríguez, E. KARPANTA: Un motor de búsqueda para la investigación experimental en Recuperación de la Información [En línea]. Salamanca: Universidad, Departamento de Informática y Automática, Facultad de Documentación. 2001. <<http://reina.usal.es>>

[GAR, 2002] García-Ojeda, J., Pérez-Alcázar, J. and Arenas, A.: Aplicación de una Metodología de Desarrollo de Sistemas Multiagente en la Diseminación Selectiva de

Información en la Web. Bucaramanga: Universidad Autónoma, Laboratorio de Cómputo Especializado. 2002.

[GIR, 2000] Giraldo, V. and Zuleta, D. La web y las bases de datos [En línea]. Antioquia, Colombia: Departamento de Ingeniería de Sistemas, Universidad. 2000. <<http://bochica.udea.edu.co/~amarin/comunica/>>

[GOB, 2005] Goble, C., Studer, R. and Finin, T. Journal of web semantics [En línea]. <[http://www.elsevier.com/wps/find/journaldescription.cws\\_home/671322/description](http://www.elsevier.com/wps/find/journaldescription.cws_home/671322/description)>

[GOM, 2002] Gomez-Perez, A. A survey on ontology tools. Madrid: Universidad Politécnica, Deliverable 1.3, Departamento de Inteligencia Artificial, Facultad de Informática. 2002. <<http://ontoweb.org/Members/huro/MyPublications/OntoWeb%20Deliverable%201.3>>

[GON, 2003] González, A. Referencia Metodológica para la Construcción de Sistemas Inteligentes, TOTEMS Software,

[GOO, 2002] Google. Google Web APIs (beta) [En línea]. <<http://www.google.com/apis/>>

[GOO, 2004] Historia de Google. Los comienzos [En línea]. <<http://google.dirson.com/historia.php>>

[GRO, 1998] Grossman, D. and Frieder, O. Information retrieval: algorithms and heuristics. Boston: Kluwer Academia Publishers. 1998

[GRU, 1993a] Gruber, T. "Toward Principles for the Design of Ontologies Used for Knowledge Sharing". Stanford: University, Stanford Knowledge Systems Laboratory. 1993.

[GRU, 1993b] Gruber, T. "A Translation Approach to Portable Ontology Specifications", Knowledge Acquisition, 199-220.

[HAR, 1992] Harman, D. Baeza-Yates, R. Information retrieval: Data Structures and Algorithms. Prentice-Hall, Englewood Cliffs, 1992. p. 363-392.

[HEN, 2004] Hendler, J. Frequently Asked Questions on W3C's Web Ontology Language (OWL) [En línea]. W3C Web Ontology Working Group. 2004.

<<http://www.w3.org/2003/08/owlfaq>>

[HER, 2003] Herrera, A. Madrid, J. Manual de Construcción de Mampostería de Concreto. Colombia: Instituto Colombiano de Productores de Cemento. 2003

[HOL, 2004] Holger, K., Ferguson, R., Noy, N. and Musen, M. The Protégé OWL Plugin: An Open Development Environment for Semantic Web Applications. Hiroshima: Third International Semantic Web Conference. ISWC. 2004.

<<http://protege.stanford.edu/plugins/owl/documentation.html>>

[IWO, 2003] IWorld. La revista de la tecnología y estrategia de negocio en Internet.: Agentes Inteligentes: cápsulas de inteligencia artificial a nuestro servicio [En línea]. España. 2003. <<http://www.idg.es/iworld/impart.asp?id=145201>>

[ISW, 2004] 3rd International Semantic Web Conference [En línea]. The Japanese Society for Artificial Intelligence and The Semantic Web Science Association. Japan. <<http://iswc2004.semanticweb.org/>>

[ISO, 2001] Requisitos ergonómicos para trabajos de oficina con pantallas de visualización de datos-PVD. [En línea]. 2001. <[http://sid.usal.es/mostrarficha.asp\\_Q\\_ID\\_E\\_7451\\_A\\_fichero\\_E\\_8.3](http://sid.usal.es/mostrarficha.asp_Q_ID_E_7451_A_fichero_E_8.3)>

[JAV, 2005] Java en Castellano. [En línea]. <<http://www.programacion.com/java/>>

[JCP, 2003] JCP. Java Community Process. Overview [En línea]. Community Development of Java Technology Specifications.

<<http://jcp.org/en/introduction/overview>>

[JON, 1997] Jones, D., Bench-Capon, T. and Visser, P.: Methodologies for Ontology Development. . Liverpool: University, Department of Computer Science. 1997.



[KRU, 2000] Krug, S. Don't Make Me Think: A Common Sense Approach to Web. New Riders. 2000.

[LIZ, 2004] Lizama, U. Perl en Español [En línea]. <<http://members.tripod.com/~MoisesRBB/1.htm>>

[LOP, 2000] López, E., García, A. and Martín, A. Tutor de Estructuras de Datos Interactivo [En línea], Granada: Universidad, Licenciatura Informática. 2000.  
<<http://decsai.ugr.es/~jfv/ed1/tedi/cdrom/docs/tablash.html> >

[LOZ, 2003] Lozano, A. Ontologías en la Web Semántica, España: Universidad de Extremadura, Área de Lenguajes y Sistemas Informáticos, Departamento de Informática. 2003.

[LOZ, 2003] Lozano-Tello, A., Macías, M., Sánchez, E. and Sosa, E. Uso de Ontologías en Páginas Web para Mejorar su Accesibilidad a Invidentes [En línea]. Extremadura: Universidad, QUERCUS Software Engineering Group. Departamento de Informática. 2003.

[MAR, 2002] Martínez, F. Propuesta y desarrollo de un modelo para la evaluación de la Recuperación de Información en Internet. Murcia: Universidad, Facultad de Ciencias de la Documentación, Tesis de doctoral. 2002.

[MIL, 2003] Milenium Network. Principales definiciones de los términos más usados en Internet. Informática [En línea]. México. 2003.  
<<http://www.informaticamilenium.com.mx/paginas/espanol/qsomos.htm>>

[MIT 2003] Massachusetts Institute of Technology (MIT). <http://www.csail.mit.edu/index.php>, <http://www.csail.mit.edu/about/about.html>.

[MON, 2003] Montero, Y. and Fernández, M. Web semántica: El papel del arquitecto de la información [En línea]. 2003. <<http://www.nosolousabilidad.com/>>

[MOR, 1999] Morales, E. Representación de Conocimiento. Mexico: Tecnológica de Monterrey (ITESM), Campus Cuernavaca, Departamento de Computación, División de Ingeniería y Ciencias. 1999.

[NOY, 2000] Noy, N., McGuinness D.: A Guide to Creating Your First Ontology  
Stanford University.

[OPP, 2000] Oppenheim, C., Morris, A., McKnight, C. and Lowley, S. "The evaluation of WWW search engines". Journal of Documentation, 56(2), 2000. p. 190-211.

[OPI, 2004] Opinamos.com empresa dedicada a la investigación de mercado digital por Internet en Latinoamérica. 2004.

<<http://opinamos.com/stats/>>

[PAV, 2004] Pavon, J.: Agentes Software. Tecnologías de la Información y la Comunicación [En línea]. Madrid: OTRI-Universidad Complutense, Sistemas informáticos y Programación, Facultad de Informática, 2004.  
<[http://www.ucm.es/info/otri/complutecno/fichas/tec\\_jpavon1.htm](http://www.ucm.es/info/otri/complutecno/fichas/tec_jpavon1.htm)>

[PEM, 2000] Pemberton, D., Rodden, T. and Procter, R. GroupMark: A WWW Recommender System Combining Collaborative and Information Filtering. Florence: 6th ERCIM Workshop "User Interfaces for All", CNR-IROE.

[PER, 2000] Perez-Carballo, J. and Strzalkowski, T. Natural language information retrieval: progress report". Information Processing and Management 36. 2000. p.155-178.

[RED, 2004] Redondo, A., Montes, M. and Romero, M. Ontologías [En línea]. Madrid: Universidad Carlos III. 2004.  
<<http://excatedra.info/enciclopedia/prueba.php?termino=Ontologias>>

[RAI, 2004] Raisa, S., Simón, A., Valdéz, R., Fernandez, F., Rosete, A., Moreno, M., Leyva., E. and Pina, J. Las ontologías en la representación del conocimiento. La Habana: Instituto Superior Politécnico "José Antonio Echeverría" (CUJAE), Centro de Estudios de Ingeniería de Sistemas (CEIS). 2004.

[RUB, 2005] Rubio, D. Fundador y Consultor Principal de OsmosisLatina.com: México: portal enfocado en software libre y código abierto. 2005.

[SAI, 2004] Saiz, F. and Castells, P. La Web semántica: Tecnologías y aplicaciones. Madrid: Universidad Autónoma, Escuela Politécnica Superior. Curso 2003 - 2004.

[SAL 1968] Salton, G. Automatic Information Organization and Retrieval New York: Mc Graw-Hill Computer Series, 1968.

[SAL, 1983], Salton, G. and McGill, M.J. Introduction to Modern Information Retrieval. New York: Mc Graw-Hill Computer Series, 1983.

[SAL, 1988] Salton, G. Buckley, C. "Term-Wighting Approaches in Automatic Text Retrieval". Information Processing and Management, 24(5), 1988. p. 513-523.

[SAU, 2002] Sauri, E., Grela, L. and Sellés, A. Ontologías en documentación [En línea]. Valencia: Universidad Cardenal Herrera-CEU. 2002.  
<<http://personales.upv.es/ccarrasc/doc/2001-2002/Ontologias/inicio.htm>>

[SDV, 2005a] SDV. Sun Developer Network.: JavaServer Pages Overview [En línea]. Sun Microsystems. Inc.  
<<http://java.sun.com/products/jsp/overview.html>>

[SDV, 2005b] SDV. Sun Developer Network. Java Servlet Technology Overview [En línea]. Sun Microsystems. Inc. 2005.  
<<http://java.sun.com/products/servlet/overview.html>>

[TIL, 2003] JADE Board. JADE: Java Agent Development Framework [En línea]. Italia: Telecom Italia Lab. 2003. <<http://jade.tilab.com/home-who.htm>>

[TOR, 1998] Torres-Moreno, J. INTERNET-2: Las nuevas redes del futuro. Montreal: Ecole Polytechnique. Département de Génie informatique.

[VEN, 2004] Vendetti, J., Crubézy, M., Dameron, O., Fergerson, R., Knublauch, H., Musen, M., Noy, N., Rubin, D. and Tu, S. Protégé ontology editor and knowledge-base framework [En línea]. Stanford: University, School of Medicine, Stanford Medical Informatics. 2004. <<http://protege.stanford.edu/index.html>>

[WAR, 1985] Ward-Harrison, A. Historical InterBase [En línea]. <<http://support.borland.com/entry.jspa?categoryID=154&externalID=2263>>

[WOO, 1999] Wooldridge, M., Jennings, N. and Kinny, D. The Gaia Methodology for Agent-Oriented Analysis and Design. Liverpool: University, Department of Computer Science. 1999.

[W3C, 1999] W3C: World Wide Web Consortium.: XML Activity [En línea]. <<http://www.w3.org/XML/>, <http://www.w3.org/XML/Core/#Publications>>

[W3C, 2001] W3C: World Wide Web Consortium.: Semantic Web Activity. <<http://www.w3.org/2001/sw/Activity>>

[W3C, 2004a] Web Ontology Working Group. OWL Web Ontology Language Overview. W3C Recommendation 10 February 2004. <<http://www.w3.org/TR/2004/REC-owl-features-20040210>>

[W3C, 2004b] W3C. Semantic Web Activity. Resource Description Framework (RDF) [En línea], <<http://www.w3.org/RDF/>>

[W3C, 2005] World Wide Web Consortium. Leading the Web to Its Full Potential....[En línea]. <<http://www.w3.org/> <http://www.w3.org/People/Berners-Lee/>>

[ZAZ, 2002] Zazo, A., Figuerola, C., Berrocal, J. and Gómez, R. Recuperación de información utilizando el modelo vectorial. Salamanca: Universidad, participación en el taller CLEF-2001, Departamento de Informática y Automática, 2002.