

PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ

FACULTAD DE CIENCIAS E INGENIERÍA



PONTIFICIA
UNIVERSIDAD
CATÓLICA
DEL PERÚ

**DESARROLLO DE UNA HERRAMIENTA PARA LA
RECUPERACIÓN DE VIDEOS EN EL DOMINIO DE LAS ARTES
MARCIALES MIXTAS**

Tesis para optar el Título de **Ingeniero Informático**, que presenta el bachiller:

James Junior Chirinos Pinedo

ASESOR: Dr. Héctor Andrés Melgar Sasieta

Lima, Febrero del 2015

RESUMEN

Con el continuo crecimiento de la subida de archivos multimedia a la Web, páginas conocidas como Youtube. Vimeo, entre otras; han estado realizando el continuo mejoramiento de sus Sistemas de Recuperación. De esta manera la cantidad de videos relevantes para el usuario no se verá afectada en la respuesta dada a este. Otros factores como el tipo de búsqueda (Sintáctica) realizada por estas páginas Web conforman esta problemática.

Por esta razón, la solución que se plantea viene en base a la búsqueda Semántica, dado que al ser una búsqueda orientada a las relaciones entre conceptos se espera una mayor eficiencia que la obtenida por la búsqueda Sintáctica la cual se basa en que las palabras usadas en la búsqueda sean las mismas que se colocaron como etiquetas al archivo multimedia.

Para llevar a cabo esta solución se realizó primero la Ontología, la cual posee los conocimientos que necesita la máquina acerca del dominio que ente caso son las Artes Marciales Mixtas. Para esto se debe tener un gran conocimiento del deporte, para reunir y relacionar los conceptos que lo conforman. Luego, se planificó como iba a actuar la herramienta, dado que hay varias opciones que se pueden realizar, con lo cual se decidió que el proceso comenzaría con un usuario haciendo una búsqueda textual, la cual nos permitirá conocer a que concepto del dominio pertenece y así a través de las relaciones de la Ontología encontrar otros conceptos relevantes; para luego consultarle a la base de datos cuales son los videos enlazados a estos conceptos.

Al tener ya el proceso de funcionamiento, se realizaron tanto la base de datos que contendría en cada fila el URL de un video y la URI de un concepto (individuo) de la Ontología; de esta manera, se sabe qué contiene cada video. Para saber exactamente si mostraba todo el contenido del concepto se usaron *flags*.

Finalmente se realizó el algoritmo que permite la búsqueda Semántica y se realizaron inferencias para búsquedas más específicas en el dominio. Teniendo la herramienta finalizada se realizaron pruebas de eficiencia usando 50 videos para lo cual se hallaron tanto la precisión como el *recall* de la herramienta al realizarse diversas consultas. Para esto se realizaron varias consultas para abarcar la mayoría de conceptos del dominio y de esta manera tener resultados más confiable.

FACULTAD DE
**CIENCIAS E
 INGENIERÍA**
 ESPECIALIDAD DE
 INGENIERÍA INFORMÁTICA

 PONTIFICIA
**UNIVERSIDAD
 CATÓLICA**
 DEL PERÚ

TEMA DE TESIS PARA OPTAR EL TÍTULO DE INGENIERO INFORMÁTICO

TÍTULO: Desarrollo de una Herramienta para la recuperación de videos en el dominio de las Artes Marciales Mixtas
ÁREA: Ciencia de la Computación
PROPONENTE: Dr. Héctor Andrés Melgar Sasieta
ASESOR: Dr. Héctor Andrés Melgar Sasieta
ALUMNO: James Junior Chirinos Pinedo
CÓDIGO: 20077241
TEMA N°: SSY
FECHA: 15/12/2014


DESCRIPCIÓN

En la actualidad dada la gran cantidad de contenido multimedia que se concentran en los portales Web como *Youtube* e *Vimeo* y el uso de etiquetas colocadas manualmente por los usuarios para la descripción del contenido, ha ido mermando la precisión de los sistemas de recuperación de videos, al hacer uso de búsqueda sintáctica. En este contexto, la recuperación de videos del dominio de las Artes Marciales Mixtas se va haciendo más complicada.

En este proyecto de fin de carrera se plantea el uso de la Web Semántica el cual permitirá afrontar este problema. El uso de la Web Semántica se da con el fin de realizar búsquedas más precisas en el dominio específico que se ha definido para de esta forma el usuario encuentre contenido multimedia que sea relevante, es decir de interés.

OBJETIVO GENERAL

Desarrollar una herramienta en el ámbito de la Web Semántica que soporte la recuperación de videos en el dominio de las Artes Marciales Mixtas.

OBJETIVOS ESPECÍFICOS

- Desarrollar un componente de software que permita enriquecer la información relacionada a los videos con elementos del dominio de las Artes Marciales Mixtas.
- Desarrollar un mecanismo que soporte la recuperación de videos en el dominio de las Artes Marciales Mixtas.
- Desarrollar un mecanismo que permita la recuperación de videos relacionados a un asunto en el dominio de las Artes Marciales Mixtas.

ALCANCE

En este proyecto de fin de carrera se desarrollará una herramienta en el ámbito de la Web Semántica para la recuperación de videos en el dominio de las Artes Marciales Mixtas; el cual pertenece al área de Ciencias de la Computación. Se realizará tanto la anotación de videos como la recuperación de los mismos en el dominio de las Artes Marciales Mixtas. Se realizará la anotación al URL de cada video, el cual se encontrará en páginas Web conocidas como lo

 Av. Universitaria 1801
 San Miguel, Lima - Perú

 Apartado Postal 1761
 Lima 100 - Perú

 Teléfono:
 (511) 526 2000 Anexo 4601





FACULTAD DE
CIENCIAS E
INGENIERÍA
ESPECIALIDAD DE
INGENIERÍA INFORMÁTICA



PONTIFICIA
UNIVERSIDAD
CATÓLICA
DEL PERÚ

son *Youtube* e *Vimeo* en el dominio de las Artes Marciales Mixtas. Luego se indexará a cada recurso los términos más importantes que se encuentren, para la posterior recuperación.

Por otra parte, se ha visto que hay una gran variedad de empresas que realizan eventos de Artes Marciales Mixtas, pero están asociadas a diferentes comisiones; en otras palabras, usan diferentes reglas lo cual significaría el uso de diferentes ontologías. Por esta razón se eligió la empresa UFC (*Ultimate Fighting Championship*), la cual realiza eventos principales con el nombre UFC seguido por una numeración y otros eventos como *UFC Fight Night*. Se eligió la empresa UFC por ser considerada la empresa número 1 en eventos de Artes Marciales Mixtas; por consiguiente, la más conocida a nivel mundial.

Máximo: 100 páginas

Av. Universitaria 1801
San Miguel, Lima - Perú

Apartado Postal 1761
Lima 100 - Perú

Teléfono:
(511) 626 2000 Anexo 4801

DEDICATORIA

Este proyecto de fin de carrera lo quiero dedicar a mis padres James y Mary; así como también a mi hermano Warren, mis abuelos Honorio y Dominica, mis tías Rosario y Doris y a mi primo Patrick; por darme las fuerzas para lograr cumplir esta meta. Además también quiero agradecer a mis amigos y a mi asesor por los alientos y el apoyo dado.

Tabla de contenido

CAPÍTULO 1	1
1 PROBLEMÁTICA	1
1.1 OBJETIVO GENERAL	4
1.2 OBJETIVOS ESPECÍFICOS	4
1.3 RESULTADOS ESPERADOS	4
2 HERRAMIENTAS, MÉTODOS, METODOLOGÍAS Y PROCEDIMIENTOS	5
2.1 INTRODUCCIÓN	5
2.2 HERRAMIENTA	5
2.2.1 OWL	5
2.2.2 PROTÉGÉ	6
2.2.3 JENA	6
2.2.4 LUCENE	6
2.2.5 MYSQL COMMUNITY EDITION	6
2.3 MÉTODOS	7
2.3.1 PRECISIÓN Y RECALL	7
2.4 METODOLOGÍAS	7
2.4.1 COMMONKADS	7
2.4.2 METODOLOGÍA 101 DE DESARROLLO DE ONTOLOGÍAS	8
3 ALCANCE	9
4 JUSTIFICATIVA	9
CAPÍTULO 2	11
5. MARCO CONCEPTUAL	11
5.1 LA TECNOLOGÍA DE LA WEB SEMÁNTICA	11
5.1.1 WEB SEMÁNTICA	11
5.1.2 ONTOLOGÍAS	12
5.1.3 ANOTACIONES	12

5.1.4 ANOTACIONES DIGITALES	13
5.1.5 METADATOS	13
5.2 EL CAMPO DE RECUPERACIÓN DE INFORMACIÓN	15
5.2.1 RECUPERACIÓN DE INFORMACIÓN	15
5.2.2 INDEXACIÓN	15
5.2.3 SISTEMAS DE RECUPERACIÓN DE INFORMACIÓN	16
5.2.4 EXPANSIÓN DE CONSULTA	17
5.3 CONCEPTOS DE ARTES MARCIALES MIXTAS	17
5.3.1 ARTES MARCIALES MIXTAS	17
5.3.2 COMISIÓN ATLÉTICA	18
5.3.3 ARBITRO Y JUECES	18
5.3.4 DOCTOR	19
5.3.5 FINALIZACIONES	19
5.4 CONCLUSIÓN	19
6. ESTADO DEL ARTE	20
6.1 FORMULACIÓN DE LA PREGUNTA	20
6.2 SELECCIÓN DE LAS FUENTES	20
6.3 TABLAS RESUMEN DEL ESTADO DEL ARTE	21
6.3.1 INVESTIGACIONES	21
6.3.2 PRODUCTOS	22
6.4 INVESTIGACIONES ACERCA DEL TEMA	22
6.4.1 WEB SEMÁNTICA PARA EL CONTENIDO BASADO EN RECUPERACIÓN DE VIDEOS	22
6.4.2 EXPLOTACIÓN DE UN CONTEXTO ADICIONAL PARA RECOMENDACIONES DE ETIQUETADO BASADO EN GRAFOS EN SISTEMAS SISTEMAS DE FOLKSONOMÍA	22
6.4.3 ONTOLOGÍAS DINÁMICAS ENRIQUECIDAS GRÁFICAMENTE PARA BIBLIOTECAS DIGITALES	23
6.4.4 DANVIDEO: UN SISTEMA MPEG-7 DE AUTORÍA Y RECUPERACIÓN DE VIDEOS DE DANZA	24
6.5 PRODUCTOS SIMILARES AL TEMA ESPECÍFICO	25
6.5.1 BLINKX	25
6.5.2 FAVIKI	26
6.5.3 FLIPBOARD	26
6.5.4 WOLFRAM ALPHA	26
6.6 CONCLUSIONES SOBRE EL ESTADO DEL ARTE	27

CAPÍTULO 3	28
7. RESULTADOS ESPERADOS DEL OBJETIVO ESPECÍFICO 1	28
7.1 ONTOLOGÍA QUE PRESENTE EL CONOCIMIENTO EN EL DOMINIO DE LAS ARTES MARCIALES MIXTAS	28
7.1.1 PASO 1: DETERMINAR EL DOMINIO Y ALCANCE DE LA ONTOLOGÍA	28
7.1.2 PASO 2: CONSIDERAR EL REÚSO DE ONTOLOGÍAS EXISTENTES	29
7.1.3 PASO 3: ENUMERAR TÉRMINOS IMPORTANTES EN LA ONTOLOGÍA	29
7.1.4 PASO 4: DEFINIR LAS CLASES Y JERARQUÍA DE CLASES	30
7.1.5 PASO 5 Y 6: DEFINIR LAS PROPIEDADES DE LAS CLASES-RANURAS Y DEFINIR LAS FACETAS DE LAS RANURAS	30
7.1.6 PASO 7: CREAR INSTANCIAS	31
7.2 ESTRUCTURA DE ANOTACIÓN SEMÁNTICA DE VIDEOS.	31
7.3 BASE DE DATOS DE VIDEOS RELACIONADOS AL DOMINIO DE LAS ARTES MARCIALES MIXTAS.	32
7.4 HERRAMIENTA QUE PERMITE AÑADIR METADATOS A LOS VIDEOS USANDO LA ONTOLOGÍA Y LA ESTRUCTURA DE ANOTACIÓN CREADA ANTERIORMENTE.	33
CAPÍTULO 4	34
8. RESULTADOS ESPERADOS DEL OBJETIVO ESPECÍFICO 2	34
8.1 HERRAMIENTA QUE PERMITE LA INDEXACIÓN DE LOS VIDEOS USANDO COMO REPRESENTACIÓN DE LOS MISMOS METADATOS	34
8.2 HERRAMIENTA QUE PERMITE RECUPERAR LOS VIDEOS USANDO BÚSQUEDA TEXTUAL	35
8.3 HERRAMIENTA QUE PERMITE RECUPERAR LOS VIDEOS USANDO BÚSQUEDA SEMÁNTICA	37
8.4 PRECISIÓN Y RECALL DE MECANISMO DE RECUPERACIÓN DE VIDEOS	38
CAPITULO 5	40
9. RESULTADOS ESPERADOS DEL OBJETIVO ESPECÍFICO 3	40
9.1 INFERENCIAS QUE PERMITEN REALIZAR BÚSQUEDAS MÁS ESPECÍFICAS	40
9.2 PRECISIÓN Y RECALL DE MECANISMO DE RECUPERACIÓN DE VIDEOS	41

CAPITULO 6	42
10. CONCLUSIONES Y TRABAJOS FUTUROS	42
10.1 CONCLUSIONES	42
10.1.1 CONCLUSIONES DEL OBJETIVO ESPECÍFICO 1	42
10.1.2 CONCLUSIONES DEL OBJETIVO ESPECÍFICO 2	43
10.1.3 CONCLUSIONES DEL OBJETIVO ESPECÍFICO 3	44
10.2 TRABAJOS FUTUROS	44
REFERENCIAS BIBLIOGRÁFICAS	46



Índice de Figuras

Figura 1: Proceso de Desarrollo de Ontología [Noy and McGuiness 2001]	8
Figura 2: Ejemplo de Ontología. Ontología de MMA. Autoría Propia	12
Figura 3: Un típico Sistema de Recuperación de Información [Rijsbergen 1979].	16
Figura 4: Encuentro de Artes Marciales Mixtas de la empresa UFC [ESPN 2011].	18
Figura 5: Resultados de las Consultas Contención [Kannan 2010].	24
Figura 6: Resultados de las Consultas Temporales [Kannan 2010].	24
Figura 7: Resultados de las Consultas Espaciales [Kannan 2010].	25
Figura 8: Resultados de las Consultas Espaciotemporales [Kannan 2010].	25
Figura 9: Jerarquía de Clases. Autoría Propia.	30
Figura 10: Ontología del dominio de las MMA. Autoría Propia.	31
Figura 11: Planteo de funcionamiento de la Herramienta. Autoría Propia.	32
Figura 12: Modelo de Base de Datos. Autoría Propia.	33
Figura 13: Archivo de texto que sirve como input a la herramienta. Autoría Propia.	33
Figura 14: Codificación realizada para la Indexación. Autoría Propia.	35
Figura 15: Codificación realizada para la Búsqueda. Autoría Propia.	36
Figura 16: Respuesta de una Búsqueda Textual. Autoría Propia.	36
Figura 17: Función que permite acceder desde un individuo a una lucha. Autoría Propia.	37
Figura 18: Herramienta que permite recuperar videos. Autoría Propia.	38
Figura 19: Codificación realizada para encontrar un peleador que realizó un <i>knockout</i> o una <i>submission</i> . Autoría Propia.	41

Índice de Tablas

Tabla 1: Mapeo de resultados esperados y las herramientas que los soportan	5
Tabla 2: Diferentes Clases de Peso con su respectiva descripción	17
Tabla 3: Cadenas generales de búsqueda avanzada para Scopus, ACM Digital Library y IEEE	20
Tabla 4: Resumen de Resultados devueltos por Bases de Datos	21
Tabla 5: Resumen de Investigaciones	21
Tabla 6: Resumen de productos	22
Tabla 7: Consultas con sus respectivas precisiones y recall	38
Tabla 8: Consultas con sus respectivas precisiones y recall	41



CAPÍTULO 1

1 Problemática

En la actualidad se suben grandes cantidades de contenido multimedia en la Web tales como videos, audios, archivos digitales entre otros, es por ello que el proceso de recuperación de estos, se realiza a través de Sistemas los cuales reciben la consulta del usuario, para luego devolver un conjunto de documentos ordenados según probabilidad de relevancia [Baeza-Yates and Ribeiro-Neto 1999]. Dado esto, páginas Web como *Youtube*, *Metacafe*, *Vimeo*, entre otras hacen uso de etiquetas para los videos, las cuales son asignadas manualmente por los usuarios [Sebastine 2009]. La asignación manual de etiquetas y la gran cantidad de contenido en video en la Web, ha traído como consecuencia la disminución de la precisión (ratio del número de documentos recuperados relevantes entre número total de documentos recuperados; siendo relevante si contiene información en la cual el usuario está interesado) por lo cual se ha tenido que realizar una mejora continua de los sistemas de recuperación de videos [Bertini 2009] [Baeza-Yates and Ribeiro-Neto 1999].

Actualmente el problema que se viene dando es que los videos no están debidamente estructurados para utilizarse en el ámbito de la Web Semántica, ya que ésta necesita de grandes cantidades de datos para lograr que la máquina comprenda la información que contiene el archivo. Esto se debe a que no presentan metadatos, en otras palabras, datos dentro de los datos que ofrezcan información acerca de lo que contiene el video [Berners-Lee, Hendler et al. 2001]. La estructura necesaria para el recurso, en el contexto de la Web Semántica, se logra gracias a herramientas como XML y RDF, las cuales proveen de estructura y significado a los recursos, por medio de anotaciones semánticas para el registro de los metadatos en los recursos [Sebastine 2009]. En buscadores de videos como *Youtube*, *Vimeo*, *Dailymotion* se hace uso de etiquetas asignadas por los usuarios, las cuales no representan completamente la información que contiene el video, esto se observa claramente en estos buscadores al tratar de subir un video, los usuarios de estas páginas Web tienen la posibilidad de colocarle el tipo de video que por ejemplo en *Youtube*, podría ser Deportes, Entretenimiento u otros, y etiquetas situadas por teclado, mientras sugiere otras etiquetas según lo que se va colocando. Es así como para videos en el dominio de las Artes Marciales Mixtas (MMA), el cual es un deporte de contacto en el cual los peleadores hacen uso de varias disciplinas como karate, judo, box, entre otros al colocar las etiquetas de esta forma se puede llegar a omitir información sobre que luchadores son los que aparecen en el video, o que empresa está organizando dicho

evento, fechas, momento del *knockout* (KO) o si en caso sucedió un sometimiento (lo cual se da cuando un peleador logra hacer rendir o desmaya con una estrangulación a su contrincante) en que *round* sucedió ésta, nombre del juez de la pelea, registro de resultados de peleas anteriores de los luchadores previas a la del video, etcétera [UFC SPORT].

Otro problema se da al realizar búsquedas con lo anteriormente expuesto acerca de las etiquetas colocadas por los usuarios y la creciente cantidad de videos subidos en la actualidad, es la disminución de la precisión. Dado que, al realizar una búsqueda, se tendrán resultados con cada vez mayor incremento de documentos no relevantes; dicho de otra manera, la precisión se verá mermada. Es acá donde la Web Semántica con el uso de metadatos asociados con estos archivos; ontologías las cuales permitirán expresar los conceptos, atributos y relaciones entre conceptos del dominio de interés; además del empleo de anotación semántica para asociar los recursos con las ontologías; nos permitirá tener mayor precisión [Bertini 2009]. Este problema en el dominio de las Artes Marciales Mixtas conlleva a que no se encuentren muchas veces lo que se desea; por ejemplo al realizar la búsqueda en *Youtube* “MMA Peruanos en el Deporte” al avanzar a la segunda página esta devuelve videos que no guardan relación con “MMA”, pero si con las palabras “Peruanos” y “Deportes”.

Un problema adicional, es la dificultad de poder recuperar videos a partir de un término del dominio. Este problema también se puede combatir con uno de los conceptos previamente mencionados siendo uno de éstos la anotación semántica, ya que de esta manera permitirá lograr este tipo de búsquedas en conjunto con la indexación (asignación términos a los documentos) y la recuperación de información (procesos que intervienen en la representación, el almacenamiento, la búsqueda y hallazgo de información relevante) [Bertini 2009]. La expansión de consulta con retroalimentación es una técnica efectiva para agregar palabras a una consulta, dando así mejores resultados en la recuperación, ya que permitirá encontrar lo más relevante acerca del asunto del dominio deseado, pudiendo esta también ser usada para resolver este problema en conjunto con lo anterior [Mitra, Singhal et al. 1998]. Este problema poniéndolo en función del dominio que son las Artes Marciales Mixtas, se da por ejemplo al querer encontrar las “peleas en las cuales se dio un *Knockout* en el segundo *round*”, ya que colocando esta frase en *Youtube* se encuentra una gran cantidad de videos que no son relevantes.

Por otro lado está el problema del incremento en el tiempo de búsqueda, ya que al darse los problemas anteriormente mencionados se terminan recuperando videos con menor eficiencia, aumentando así el tiempo que demorara el usuario en encontrar un video con contenido relevante; para resolver esto se puede hacer uso de algoritmos que ordenen los videos según su relevancia [Abel 2008]. Un ejemplo de este tipo de algoritmos seria el *FolkRank* el cual está basado en el *PageRank* de *Google*, el cual tiene como finalidad formar un ranking según la relevancia de los documentos [Abel 2008]. Además de usar algoritmos, se puede hacer uso de la indexación la cual permite indexar términos según la frecuencia de aparición que tengan en el recurso [Salton 1981]. Llevando este problema al dominio de las Artes Marciales Mixtas se ve que este se da comúnmente, ya que al buscar algo acerca del Artes Marciales Mixtas, como por ejemplo “mma anaconda choke primer round” uno se puede pasar revisando videos uno por uno hasta encontrar alguno que le parezca relevante, ya que dentro de los resultados en las primeras páginas se pueden encontrar las formas de cómo realizar esta técnica y no acerca de peleas donde se dio esta técnica en el primer round, lo que sí se encuentra pero en resultado posteriores. Por esta razón, al no encontrar el usuario entre los primeros resultados los videos que a su juicio son relevantes se incrementará el tiempo de búsqueda.

Otro problema es la dificultad para encontrar la mayor cantidad de recursos de forma precisa, y con un aumento en la cantidad de documentos relevantes recuperados sobre el número de documentos relevantes devueltos y no devueltos, siendo esto último llamado *recall* [Rijsbergen 1979]. En la actualidad dado que se sube una gran cantidad de videos a la Web se ha reducido la precisión de las búsquedas así como también se ha reducido el *recall* [Bertini 2009]. Este problema se puede enfrentar mediante el uso de un sistema de recuperación de información y la expansión de consulta, siendo el primero de estos el cual le da una mayor facilidad de encontrar términos relevantes en los recursos y el segundo mejorando la calidad de la recuperación [Singhal 2001]. Este problema siendo observado en el dominio, dada la naturaleza de los buscadores actuales, y por lo visto anteriormente en el problema de la búsqueda por un asunto del dominio, se ve representado por la cantidad de videos recuperados en los buscadores como *Youtube* de los cuales solo pocos son relevantes y encuentran una gran cantidad de videos que no lo son.

Se puede concluir que el problema central es la dificultad de poder recuperar videos en el dominio de las Artes Marciales Mixtas. Por esta razón, nos preguntamos lo siguiente **¿De qué manera se pueden recuperar videos relevantes en el dominio**

de las Artes Marciales Mixtas en el ámbito de la Web? La respuesta a esta pregunta nos llevará a nuestro objetivo general.

1.1 Objetivo general

Desarrollar una herramienta en el ámbito de la Web Semántica que soporte la recuperación de videos en el dominio de las Artes Marciales Mixtas.

1.2 Objetivos específicos

Objetivo 1: Desarrollar un componente de software que permita enriquecer la información relacionada a los videos con elementos del dominio de las Artes Marciales Mixtas.

Objetivo 2: Desarrollar un mecanismo que soporte la recuperación de videos en el dominio de las Artes Marciales Mixtas.

Objetivo 3: Desarrollar un mecanismo que permita la recuperación de videos relacionados a un asunto en el dominio de las Artes Marciales Mixtas.

1.3 Resultados esperados

Resultados del Objetivo 1:

1. Ontología que represente el conocimiento en el dominio de las Artes Marciales Mixtas.
2. Estructura de anotación semántica de videos.
3. Base de datos de videos relacionados al dominio de las Artes Marciales Mixtas.
4. Herramienta que permite añadir metadatos a los videos usando la ontología y la estructura de anotación creada anteriormente.

Resultados del Objetivo 2:

5. Herramienta que permite la indexación de los videos usando como representación de los mismos metadatos.
6. Herramienta que permite recuperar los videos usando búsqueda textual.
7. Herramienta que permite recuperar los videos usando búsqueda semántica.
8. Precisión y *Recall* de mecanismo de recuperación de videos.

Resultados del Objetivo 3:

9. Inferencias que permiten realizar búsquedas más específicas
10. Precisión y *Recall* de mecanismo de recuperación de videos.

2 Herramientas, métodos, metodologías y procedimientos

2.1 Introducción

En este punto se pasarán a señalar y describir las herramientas, métodos, metodologías y procedimientos que se usarán para este desarrollo. En la Tabla 1 se muestra el mapeo de las herramientas usadas con los resultados esperados que serán realizados con apoyo de estas.

Tabla 1: Mapeo de resultados esperados y las herramientas que los soportan

Resultados esperados	Herramientas a usarse
1. Ontología que represente el conocimiento en el dominio de las Artes Marciales Mixtas.	Protégé, OWL, Ontology Development 101 y CommonKads
2. Estructura de anotación semántica de videos.	CommonKads
3. Base de datos de videos relacionados al dominio de las Artes Marciales Mixtas.	MySQL y CommonKads
4. Herramienta que permite añadir metadatos a los videos usando la ontología y la estructura de anotación creada anteriormente.	MySQL y CommonKads
5. Herramienta que permite la indexación de los videos usando como representación de los mismos metadatos.	Lucene, Jena, MySQL y CommonKads
6. Herramienta que permite recuperar los videos usando búsqueda textual.	Lucene y CommonKads
7. Herramienta que permite recuperar los videos usando búsqueda semántica.	Jena, MySQL y CommonKads
8. Precisión y <i>Recall</i> de mecanismo de recuperación de videos.	Precisión y Recall
9. Inferencias que permiten realizar búsquedas más específicas	MySQL, Jena, y CommonKads
10. Precisión y <i>Recall</i> de mecanismo de recuperación de videos.	Precisión y Recall

2.2 Herramienta

2.2.1 OWL

Ontology Web Language (OWL) es un lenguaje de Web Semántica diseñada para representar conocimiento rico y complejo sobre cosas, grupos de cosas y relaciones entre cosas. OWL es un lenguaje basado en lógica computacional, ya que puede ser aprovechado por programas de computadora para por ejemplo verificar la consistencia del conocimiento o para hacer implícito conocimiento explícito. Los documentos de OWL son conocidos como ontologías, las cuales pueden ser publicadas en la *World Wide Web* y pueden referenciar o ser referenciadas por otras ontologías de OWL [OWL W3C]. OWL 2 es un tipo de OWL que provee de clases, propiedades, individuos, y valores de datos que son almacenados como documentos de Web Semántica [OWL2 W3C]. Al tener que desarrollarse ontologías, esto se realizará con el lenguaje de Web Semántica OWL 2 el cual es un lenguaje diseñado para crear ontologías.

2.2.2 *Protégé*

Es un ambiente gratuito para el modelado de ontologías. Actualmente existe una versión Web y una versión de escritorio que puede ser usada en entornos Windows y Linux. Protégé provee a su creciente comunidad de usuarios con una serie de herramientas para construir modelos de dominio y aplicaciones basadas en conocimiento con ontologías [Protege Stanford]. En su núcleo, implementa un conjunto de estructuras de modelado de conocimiento enriquecidas y acciones que soportan la creación, visualización de ontologías en varios formatos de representación [Protege Stanford].

2.2.3 *Jena*

Es un *framework* de Java para construir aplicaciones en la Web Semántica [Jena Apache]. Este provee de librerías de Java para ayudar a los desarrolladores a elaborar código que maneje RDF, RDFS, RDFa, OWL y SPARQL alineado con las recomendaciones W3C publicadas [Jena Apache]. Jena incluye un motor de inferencia basado en reglas para lograr el razonamiento basado en ontologías OWL y RDFS, y una variedad de estrategias de almacenamiento para almacenar triplas RDF en memoria o en disco [Jena Apache]. Dada la necesidad de construir una aplicación en la Web semántica, se da la necesidad de usar una herramienta que provea un ambiente de programación para OWL, método posteriormente mencionado, para lo cual se ha seleccionado el *framework* Jena.

2.2.4 *Lucene*

Lucene es una librería caracterizada por ser un motor de búsqueda de texto, con todas sus funciones escritas completamente en Java. Es una tecnología adecuada para casi cualquier aplicación que requiera la búsqueda de texto, especialmente multiplataforma [Lucene Core]. Dado la necesidad de recuperar información, se hará uso de la librería Lucene.

2.2.5 *MySQL Community Edition*

MySQL Community Edition es la versión de la base de datos de código libre gratuita. Está disponible bajo licencia GPL y es soportada por una comunidad activa de desarrolladores de código libre [MySQL COMM]. Está Disponible para 20 plataformas y sistemas operativos incluyendo *Linux, Unix, Mac y Windows*.

2.3 Métodos

2.3.1 Precisión y *Recall*

La precisión y *Recall* son los criterios estándar para la evaluación de la recuperación de la información. La Precisión se encuentra dada por la relación entre el número de documentos recuperados relevantes entre el número de documentos recuperados. El *Recall* está definido como el número de documentos recuperados relevantes, entre el número total de documentos relevantes en la colección; en otras palabras, de todos los videos [Ingwersen 1992]. Su uso es necesario dado que se necesita evaluar la eficiencia de la herramienta.

2.4 Metodologías

2.4.1 *CommonKADS*

CommonKADS es la metodología líder para el soporte de ingeniería del conocimiento estructurado; este ha sido gradualmente desarrollado y ha sido validada por muchas compañías y universidades en el contexto del European ESPRIT IT Programme [CommonKADS DEF]. *CommonKADS* permite detectar oportunidades y cuellos de botella en cómo se desarrollan las organizaciones, distribuyen y aplican sus recursos de conocimiento, y por lo tanto da las herramientas para la gestión del conocimiento corporativo. Además, *CommonKADS* proporciona métodos para llevar a cabo un análisis detallado de las tareas y procesos intensivos de conocimiento. Finalmente, *CommonKADS* da apoyo al desarrollo de sistemas de conocimiento que soportan partes seleccionadas del proceso del negocio [CommonKADS DEF].

CommonKADS ofrece un enfoque estructurado, el cual está basado en pensamientos o principios que se han desarrollado por la experiencia a través de los años [Schreiber 2000]. La ingeniería del conocimiento, no es una extracción desde la cabeza del experto, sino que consiste en construir diferentes modelos de aspectos del conocimiento humano [Schreiber 2000].

- El principio del nivel de conocimiento: en el modelado del conocimiento, primero nos concentramos en la estructura del conocimiento, y dejamos los detalles de programación para después [Schreiber 2000].
- El conocimiento tiene una estructura interna estable que es analizable para distinguir tipos y roles de conocimiento específico [Schreiber 2000].
- Un proyecto de conocimiento debe ser gestionado para aprender a partir de tus experiencias en un camino “espiral” controlado [Schreiber 2000].

Las ontologías representan el conocimiento, definiendo términos en el dominio y las relaciones a través de ellos [Noy and McGuiness 2001]. Por ello, son abarcadas dentro de lo que es Ingeniería de Conocimiento. En conclusión esta metodología puede ser usada en este trabajo de fin de carrera.

2.4.2 Metodología 101 de Desarrollo de Ontologías

Es una metodología para el desarrollo de ontologías, el cual posee un enfoque iterativo, el cual está conformado por 7 pasos, los cuales se deben realizar siguiendo las siguientes reglas [Noy and McGuiness 2001]:

- No hay una manera correcta de modelar un dominio- siempre hay alternativas viables. La mejor solución casi siempre depende de la aplicación que tienes en mente y las extensiones que anticipas,
- El desarrollo de ontologías es necesariamente un proceso iterativo,
- Los conceptos en la ontología deberían ser cercanos a los objetos (físicos o lógicos) y relaciones en tu dominio de interés. Tienen más probabilidades de ser sustantivos (objetos) o verbos (relaciones) en sentencias que describen tu dominio [Noy and McGuiness 2001].

Los 7 pasos de la metodología son los siguientes:

- Determinar el dominio y alcance de la ontología,
- Considerar reusar ontologías existentes,
- Enumerar términos importantes en la ontología,
- Definir las clases y la jerarquía de clases,
- Definir las propiedades de clases- ranuras,
- Definir las facetas de las ranuras,
- Crear instancias [Noy and McGuiness 2001].

La Figura 1 muestra el proceso de Desarrollo de una ontología.



Figura 1: Proceso de Desarrollo de Ontología [Noy and McGuiness 2001]

Esta metodología será necesaria para realizar nuestra ontología, ya que nos da los pasos a seguir para plasmar el conocimiento que tenemos en la ontología.

3 Alcance

En este trabajo de fin de carrera se desarrollará una herramienta en el ámbito de la Web Semántica para la recuperación de videos en el dominio de las Artes Marciales Mixtas; el cual pertenece al área de Ciencias de la Computación; en otras palabras, se realizará tanto la anotación de videos como la recuperación de los mismos en el dominio de las Artes Marciales Mixtas. Se realizará la anotación al URL de cada video, el cual se encontrará en páginas Web conocidas como lo son *Youtube*, *Vimeo* entre otras y pertenecerá al dominio de las Artes Marciales Mixtas. Luego se indexará a cada recurso los términos más importantes que se encuentren, para la posterior recuperación.

Por otra parte, se ha visto que hay una gran variedad de empresas que realizan eventos de Artes Marciales Mixtas, pero están asociadas a diferentes comisiones; en otras palabras, usan diferentes reglas lo cual significaría el uso de diferentes ontologías. Por esta razón se eligió la empresa UFC (*Ultimate Fighting Championship*), la cual realiza eventos principales con el nombre UFC seguido por una numeración y otros eventos como *UFC Fight Night*. Se eligió la empresa UFC por ser considerada la empresa número 1 en eventos de Artes Marciales Mixtas; por consiguiente, la más conocida a nivel mundial.

4 Justificativa

Este trabajo de fin de carrera trata acerca del desarrollo de una herramienta en el ámbito de la Web Semántica para la recuperación de videos en el dominio de las Artes Marciales Mixtas. Es por ello, que en este proyecto se realizará la anotación de videos, de esta manera se logrará que los videos contengan información acerca de las Artes Marciales Mixtas. Esta anotación se puede realizar de distintas formas, o al URL de los videos como se está haciendo en este caso o al archivo en sí; teniendo esta última diversas formas de realizarse. Esto se puede ver en proyectos similares dados en otros contextos donde por ejemplo se realiza anotación por capas del video, en otras palabras se realizará una anotación a cada *frame* del video [Sebastine 2009]. En conclusión este trabajo de fin de carrera mostrará una forma de las diversas que hay de anotación.

Por otra parte, este proyecto se puede desarrollar en otros dominios, habiéndose elegido en este trabajo de fin de carrera las Artes Marciales Mixtas. De esta manera siendo un deporte en crecimiento en todo el mundo, esta sería una gran forma de poder compartir conocimientos acerca de éste y de cierta manera ayudar a difundirlo en nuestro país donde aún es desconocido por la mayoría. Además, este proyecto

podría ser de utilidad a personas que recién están aprendiendo o que recién están practicando el deporte, ya que podrían realizar desde consultas para observar a los mejores peleadores del deporte, hasta para realizar búsquedas por movimientos y observar a todos los peleadores que los realizan. Algo similar a los que se puede ver en el proyecto DanVideo, el cual es una herramienta de recuperación de videos de danza en el ámbito de la Web Semántica, el cual se hizo con la finalidad de que personas que recién están aprendiendo puedan hacerlo viendo a expertos [Kannan 2010].



CAPÍTULO 2

5. Marco Conceptual

En el presente marco conceptual se presentará la definición de los términos extraídos de la investigación bibliográfica realizada, los cuales serán conceptualizados para permitir el entendimiento del problema seleccionado. El objetivo es comprender los conceptos relacionados a Web Semántica y recuperación de información, para que pueda entender de una mejor manera el tema de este trabajo de fin de carrera; además de dar a conocer lo que son las Artes Marciales Mixtas, dándole así al lector la posibilidad de conocer el dominio.

5.1 La Tecnología de la Web Semántica

5.1.1 Web Semántica

Tim Berners-Lee tuvo una visión de 2 partes para el futuro de la Web; la primera es hacer una Web más colaborativa y la segunda es hacer una Web entendible y procesable para las máquinas, con lo cual la información tendría un papel más importante del que tiene actualmente [Fensel et al. 2003] [Antoniou 2004]. La Web Semántica es una extensión de la Web actual, en donde la información tiene un significado bien definido, lo cual permite que personas y máquinas trabajen en cooperación. Para darse el funcionamiento de la Web Semántica, las máquinas deben tener acceso a colecciones estructuradas de información y conjuntos de reglas de inferencia que pueden conducir para el razonamiento automatizado [Berners-Lee, Hendler et al. 2001]. La Web Semántica otorgará estructura al significado del contenido más importante de las páginas Web, creando un entorno donde agentes de software podrán realizar fácilmente tareas para los usuarios interactuando entre paginas [Berners-Lee, Hendler et al. 2001]. La Web Semántica permite que las máquinas comprendan documentos y datos semánticos [Berners-Lee, Hendler et al. 2001].

En la Web Semántica, lo opuesto a la Web como la conocemos hoy, las máquinas podrían ser capaces de navegar, integrar y procesar información de una manera significativa [Studer, Decker et al. 2000]. Por ejemplo en este trabajo de fin de carrera se anotarán los URL de los videos; en otras palabras, estos contendrán información acerca de su contenido; de esta forma, las máquinas lograrán entender que contienen los videos.

5.1.2 Ontologías

Una ontología es una especificación explícita de una conceptualización. Una conceptualización es un resumen; en otras palabras, una vista simplificada del mundo que se desea representar para algún propósito. Todo conocimiento base, sistema basado en conocimiento, o agente de nivel de conocimiento está comprometido a alguna conceptualización, explícita o implícita [Gruber 1995] [Kiryakov, Popov et al. 2004].

Las ontologías representan el conocimiento, definiendo términos en el dominio y las relaciones a través de ellos [Noy and McGuinness 2001]. En la Figura 2 se muestra un ejemplo de una ontología que conceptualiza las Artes Marciales Mixtas donde se ven clases como *Fight*, *Event*, *Fight*, *TitleFight* entre otros.

Se dice que las ontologías son la tecnología clave para la Web Semántica, ya que proveen la estructura para facilitar la representación de los conceptos y así proveer niveles de incremento de detalles [O'Leary 1998]. Además, las ontologías nos dan la capacidad de describir y representar en una máquina, de una manera utilizable una parte de nuestros modelos mentales sobre dominios específicos [Fensel et al. 2003].

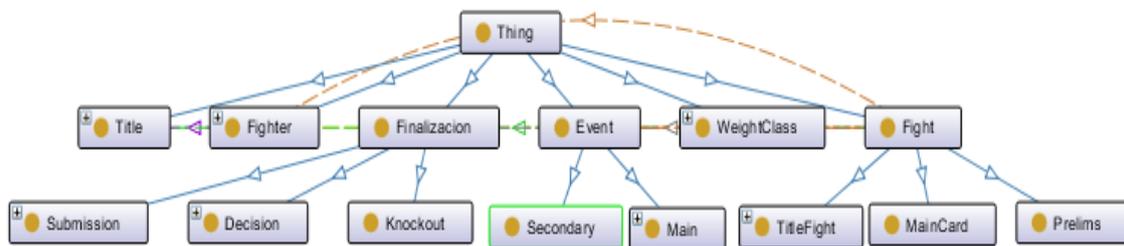


Figura 2: Ejemplo de Ontología. Ontología de MMA. Autoría Propia

5.1.3 Anotaciones

Las anotaciones se pueden definir como los comentarios, notas, explicaciones, u otros tipos de observaciones externas que están enlazados a una parte o a todo un documento. Al ser las anotaciones externas, se pueden hacer sin modificar la estructura o el contenido del documento. Las anotaciones son consideradas como metadatos ya que brindan información adicional acerca de una pieza de dato existente; además estas se distinguen por tener como propiedades distintivas: Localización física (puede ser almacenada localmente en uno o más servidores de anotaciones), Alcance (puede ser asociada con una parte o todo un documento) y tipo (tienen varios tipos funcionales como comentarios, consultas entre otros) [Meena, Kumar et al. 2004]. El siguiente ejemplo nos muestra una anotación, en la cual se

muestra la descripción de un encuentro de Artes Marciales Mixtas el cual se encuentra en el video del URL.

```

</rdf:Description>
</rdf:RDF>
<rdf:Description rdf:about="http://youtube.com/watch?v=uV0zFPnKDTs">
  <j.0:videoLength>640</j.0:videoLength>
  <dc:date>21/08/2012</dc:date>
  <dc:description>Segunda pelea dada en la UFC entre Anderson Silva
y Sonnen</dc:description>
  <dc:creator>BPJPro</dc:creator>
  <j.0:videoWidth>451</j.0:videoWidth>
  <dc:enterprise>UFC</dc:enterprise>
  <dc:nameEvent>UFC 148</dc:nameEvent>
  <dc:fighter1>Anderson Silva</dc:fighter1>
  <dc:fighter2>Chael Sonnen</dc:fighter2>
  <dc:winner>Anderson Silva</dc:winner>
  <dc:winBy>Nocaut</dc:winBy>
</rdf:Description>

```

Las personas han venido usando las anotaciones haciendo marcas que les ayuden reconocer las partes importantes de un texto, realizando esta actividad al momento de la lectura [Marshall 1997]. Con el paso a las versiones digitales de los documentos se han tenido inconvenientes al anotar, ya que se quiere trasladar lo que ya se hacía en papel a lo digital, siendo el gran problema que las anotaciones no eran tan fácil como lo eran en papel [O'Hara and Sellen 1997]. Por ello al diseñar anotaciones se debe permitir que los lectores recuperen fácilmente sus anotaciones y categorizarlas según contenido positivo o negativo [Wolfe 2000].

5.1.4 Anotaciones Digitales

Las anotaciones digitales se encuentran ligadas al archivo anotado y contienen una faceta de la información que nos informa la utilidad o no de este archivo para uno. A este enfoque se le denomina "orientado a la anotación" [Sannomiya, Amagasa et al. 2001]. Estas anotaciones poseen dos dimensiones formal e informal, de las cuales tenemos a los metadatos como los más formales, ya que siguen una estructura estándar y les es asignado valores usando nombramiento de autoridades convencionales [Marshall 1998]. La diferencia entre las anotaciones formales e informales es la interoperabilidad [Marshall 1998]. Entre las anotaciones más informales tenemos las que son del tipo de las que escribimos en un artículo de periódico [Marshall 1998].

5.1.5 Metadatos

Los metadatos son datos acerca de los datos, son una pieza secundaria de información que es independiente de alguna manera de la parte primera de información a la que se refiere. Ejemplos de metadatos incluyen esquemas, integridad

de restricciones, comentarios sobre datos, ontologías, parámetros de calidad, anotaciones, procedencia y políticas de seguridad [Srivastava and Velegrakis 2007] [Agosti and Ferro 2007] [Uren, Cimiano et al. 2006].

Los metadatos son usados para proveer alguna formalización parcial del contenido del documento como un pre-requisito para un manejo integral [Kiryakov, Popov et al. 2004]. Los metadatos pueden ser aplicados a documentos en diferentes formatos como HTML, PDF, entre otros; pero no solo puede ser aplicados a documentos, sino también a aplicaciones que se encuentren ejecutando en nuestras computadoras o que se están disponibles en la Web en forma de servicios Web. Los metadatos pueden ser expresados en un gran rango de lenguajes y vocabularios, pueden estar disponibles en diferentes formatos y pueden ser mantenidos usando diferentes herramientas [Corcho 2006].

Los metadatos pueden beneficiar a las personas porque si ellos son expresados de una forma que pueda ser leído por humanos y sus formatos y campos son conocidos, estos podrán usarlos para obtener información útil y bien estructurada sobre un contenido existente; por otra parte los metadatos ofrecen a las computadoras el significado para procesar automáticamente los contenidos anotados [Agosti and Ferro 2007].

El siguiente es un ejemplo de metadato acerca de un encuentro de Artes Marciales Mixtas entre los peleadores Anderson Silva y Sonnen.

```

<metadata>
  <rdf:RDF
    xmlns:rdf = "http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:rdfs = "http://www.w3.org/2000/01/rdf-schema#"
    xmlns:dc = "http://purl.org/dc/elements/1.1/" >
    </rdf:Description>
  </rdf:RDF>
  <rdf:Description
    rdf:about="http://youtube.com/watch?v=uV0zFPnKDTs">
    <j.0:videoLength>640</j.0:videoLength>
    <dc:date>21/08/2012</dc:date>
    <dc:description>Segunda pelea dada en la UFC entre Anderson
    Silva y Sonnen</dc:description>
    <dc:creator>BPJPro</dc:creator>
    <j.0:videoWidth>451</j.0:videoWidth>
    <dc:enterprise>UFC</dc:enterprise>
    <dc:nameEvent>UFC 148</dc:nameEvent>
    <dc:fighter1>Anderson Silva</dc:fighter1>
    <dc:fighter2>Chael Sonnen</dc:fighter2>
    <dc:winner>Anderson Silva</dc:winner>
    <dc:winBy>Nocaut</dc:winBy>
  </rdf:Description>
</rdf:RDF>
</metadata>

```

5.2 El Campo de Recuperación de Información

5.2.1 *Recuperación de Información*

La recuperación de la información se ocupa de los procesos que intervienen en la representación, el almacenamiento, la búsqueda y hallazgo de la información la cual es relevante para un requerimiento de información deseada por un usuario. Siendo relevante si contiene información en la cual el usuario está interesado [Ingwersen 1992] [Baeza-Yates and Ribeiro-Neto 1999] [Cerulo and Canfora 2004]. Una idea importante en la recuperación de la información es “la relevancia”, ya que es importante, que al recuperar información se obtengan todos los documentos relevantes; al mismo tiempo que se trata de obtener la menor cantidad de documentos no relevantes [Rijsbergen 1979].

La recuperación de información se diferencia de la recuperación de datos, en que la recuperación de información usualmente trata con texto en lenguaje natural, el cual no está siempre bien estructurado y podría ser semánticamente ambiguo; por otra parte la recuperación de datos trata con datos que tienen una estructura y semántica bien definida [Baeza-Yates and Ribeiro-Neto 1999].

5.2.2 *Indexación*

Se le llama así a la tarea de asignar los términos a los documentos individuales; por consiguiente esta tarea es crucial, ya que un error en esta podría acarrear fallas en la recuperación. Se han desarrollado diferentes técnicas de indexado, entre estas tenemos la indexación hecha manualmente por indexadores entrenados o expertos y el sistema de indexación de texto completo el cual es automático como las primeras estrategias.

Luego en los siguientes años se desarrollaron otras técnicas como la de importancia del término y frecuencia del término, donde para realizar la indexación se toma en cuenta la frecuencia de un término; por esta razón se dan dos reglas, una de ellas es la frecuencia en el documento de un término, el cual es la frecuencia de ocurrencia total dada en un documento, y la otra es la frecuencia en una colección, que es el número de documentos en una colección en la cual el término aparece. La especificidad del término y la discriminación del término, el uso de relación de términos, la interacción con el sistema de usuario, y un modelo de indexación automática son otras técnicas de indexación que surgieron a través de los años [Salton 1981].

5.2.3 Sistemas de Recuperación de Información

Un Sistema de Recuperación de Información posee 3 componentes los cuales son los siguientes: entrada, procesador y salida, tal como se observa en la Figura 3. En la entrada se encuentra el principal problema el cual es conseguir una representación de cada documento y consulta adecuado para el uso en una computadora. El procesador es la parte encargada del proceso de recuperación, este proceso puede implicar la estructuración de la información de una manera apropiada. La salida es usualmente un conjunto de citas o números de documento [Rijsbergen 1979].

La eficiencia de un sistema de recuperación es medida usualmente en términos de recursos de computadoras usados; es difícil medir la eficiencia de una manera que sea independiente a una máquina. La eficacia se mide por la precisión (ratio del número de documentos recuperados relevantes al número total de documentos recuperados) y el recall (ratio del número de documentos relevantes devueltos al número de documentos relevantes tanto devueltos como no devueltos) [Rijsbergen 1979] [Salton 1981].

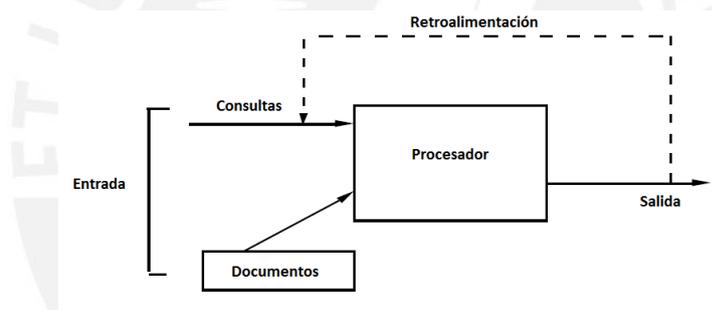


Figura 3: Un típico Sistema de Recuperación de Información [Rijsbergen 1979].

En un sistema de recuperación los documentos son representados a través de un conjunto de términos indexados o palabras clave; tales palabras claves son extraídas directamente desde el texto del documento o podrían ser especificados por una persona [Baeza-Yates and Ribeiro-Neto 1999] [Salton 1981]. Por ejemplo se tendría una alta precisión si de 20 videos recuperados 18 son relevantes para el usuario; además de un recall alto si no se recuperaron 2 videos relevantes para el usuario.

El proceso de recuperación de información puede iniciar una vez que la base de datos de documentos esta indexada; primeramente el usuario especifica sus necesidades las cuales son analizadas y transformadas por las mismas operaciones textuales aplicadas al texto. Entonces, las operaciones de consulta pueden ser aplicadas antes de la consulta actual, lo cual provee una representación del sistema para que la necesidad del usuario sea generada. La consulta luego es procesada para la recuperación de documentos, la cual se realiza de una manera rápida gracias a la

estructura indexada previamente construida. Antes de enviarse al usuario, los documentos recuperados son categorizados de acuerdo a la probabilidad de relevancia. Finalmente, el usuario revisa los documentos recuperados en búsqueda de información útil [Baeza-Yates and Ribeiro-Neto 1999].

5.2.4 Expansión de Consulta

Es una técnica efectiva usada para agregar palabras útiles a una consulta, pero desafortunadamente rara vez los usuarios casuales proveen un sistema con el juicio de relevancia necesario para mejorar las consultas; por lo cual, se tiene que hacer uso de *ad hoc* o retroalimentación ciega. La retroalimentación ciega no requiere de la entrada del usuario, ya que recupera un pequeño conjunto de documentos con la consulta original, los cuales son tomados como relevantes sin ninguna intervención del usuario y usados para expandir la consulta; este método tiene la desventaja de que la mayoría de documentos del pequeño conjunto recuperado, no sean relevantes [Mitra, Singhal et al. 1998] [Singhal 2001] [Bhagal, Macfarlane et al. 2007]. Por ejemplo, se podría colocar como consulta “Anderson Silva” y “Sonnen”, agregándose posiblemente por expansión “UFC 148”, ya que es un evento donde sucedió el encuentro entre los 2.

5.3 Conceptos de Artes Marciales Mixtas

5.3.1 Artes Marciales Mixtas

Las Artes Marciales Mixtas es un deporte de *full contact*, en el cual los peleadores poseen entrenamiento en varias disciplinas, para que de esta forma tengan un balance y sean luchadores completos, pudiendo pelear tanto en el suelo como de pie [UFC SPORT]. La duración de los combates es de 3 rounds siempre que sea una pelea que no sea por campeonato; en caso contrario, se dan 5 rounds de máximo 5 minutos; teniendo siempre un minuto de descanso entre rounds. En la Tabla 2 se observan las categorías por peso que se presentan en las Artes Marciales Mixtas.

Tabla 2: Diferentes Clases de Peso con su respectiva descripción

Peso	Descripción
Peso mosca	125 libras o menos (56,7 kg o menos)
Peso gallo	de 125 a 135 libras (de 56,7 a 61,2 kg)
Peso pluma	de 135 a 145 libras (de 61,2 a 65,7 kg)
Peso ligero	de 145 a 155 libras (de 65,7 a 70,3 kg)
Peso welter	de 155 a 170 libras (de 70,3 a 77,1 kg)
Peso medio	de 170 a 185 libras (de 77,1 a 83,9 kg)
Peso semicompleto	de 185 a 205 libras (de 83,9 a 92,9 kg)
Peso completo	de 205 a 265 libras (92,9 a 120,2 kg)
Peso supercompleto	más de 265 libras (más de 120,2 kg)

En la actualidad existen varias empresas, las cuales realizan eventos que difunden este deporte y que siguen las reglas unificadas y otras importantes regulaciones de Artes Marciales Mixtas, entre ellas tenemos como las más importantes: UFC, Bellator, Pride, Strikeforce. La más conocida de las mencionadas anteriormente es la UFC, la cual comenzó como una competencia que buscaba la mejor arte marcial, siendo así como con el tiempo los peleadores se dieron cuenta que tenían que tener conocimiento de más de una disciplina para poder competir entre los mejores; esta empresa realiza mensualmente eventos correlativos en diferentes lugares del mundo [UFC RULES]. En la Figura 4 se puede observar un encuentro de Artes Marciales Mixtas en el cual se enfrentaron Silva y Sonnen por el UFC 117.

5.3.2 Comisión Atlética

El mandato de toda comisión es asegurar que los atletas estén protegidos y velar por la seguridad del atleta que es de suma importancia. Las reglas que impone la comisión atlética son un intento de proporcionar la gobernabilidad y la claridad a la competencia de Artes Marciales Mixtas, y brindar a todos los atletas protección [UNIF RULES].

5.3.3 Arbitro y Jueces

El árbitro, el cual es designado por la comisión, se encuentra en el ring al igual que los peleadores, desde donde observa el encuentro en donde de acuerdo a las situaciones que se den, tendrá que actuar según el reglamento establecido por la comisión. El árbitro tiene poder de penalizar a alguna de los peleadores, de parar la pelea, de notificar al luchador que cometió una falta y de separar a los contendientes. Los 3 jueces son los encargados de puntuar los encuentros, cada juez tiene la potestad de otorgar 10 puntos como máximo. Al final del encuentro designan al ganador o el empate según el puntaje obtenido por los peleadores [UNIF RULES].



Figura 4: Encuentro de Artes Marciales Mixtas de la empresa UFC [ESPN 2011].

5.3.4 *Doctor*

El doctor tiene que ingresar al ring en cuanto el árbitro señale una falta en contra del peleador de su esquina. Al entrar al ring el doctor debe demorar menos de 5 minutos en analizar al luchador, ya que si se excede este tiempo la pelea no se podrá continuar [UNIF RULES].

5.3.5 *Finalizaciones*

Existen 3 maneras de acabar un combate las cuales son Nocaout, Sometimiento y Decisión. El Nocaout se da cuando un peleador queda inconsciente, noqueado, desorientado o imposibilitado de defenderse inteligentemente, dado un golpe certero del contrincante. El Sometimiento se realiza cuando un peleador consigue que el oponente se rinda o desmaye con una estrangulación (presión en el cuello para evitar el flujo de sangre al cerebro o el flujo de aire a los pulmones) o llave (extender articulación del oponente al punto máximo, causando dolor extremo). Por último, la decisión se da cuando se han acabado todo los *rounds* sin que ningún peleador dé un paso atrás, para esto existen 3 jueces quienes dan puntuación por cada *round* siendo el ganador de un *round* acreedor a 10 puntos y el oponente a 9 puntos o menos, por último el ganador de la pelea es quien consiguió más puntos de los 3 jueces. Existen 4 formas de decisión siendo estas Decisión Unánime (tres jueces están de acuerdo en el ganador), Decisión Dividida (dos jueces eligen a un mismo ganador y el tercer juez elige al otro peleador), Decisión Mayoritaria (dos jueces eligen a un mismo ganador y el tercer juez vota empate), Empate Mayoritario (dos jueces votan empate y el tercer juez elige un ganador) [UFC GANAR].

5.4 *Conclusión*

En conclusión, con los conceptos vistos de Web Semántica se permitirá la comprensión de cuán importante son para realizar este tipo de recuperación de videos al tener la información de los videos disponible. Además con los conceptos mostrados de recuperación de videos se pudo observar las distintas técnicas de indexación, la recuperación de información, la expansión de consulta y conceptos importantes como precisión y *recall*; siendo estos 2 últimos los que permitirán medir la eficiencia de la herramienta y los anteriores los que nos ayudaron a comprender el cómo funciona la recuperación. Por último si el lector no tiene conocimientos acerca del dominio que son las Artes Marciales Mixtas, se presentaron conceptos que podrán ayudarlo a entender cómo se da este deporte.

6. Estado del arte

En esta sección se darán a conocer proyectos, investigaciones y productos relacionados tanto a la Web Semántica como a la recuperación de videos, esto se logrará gracias al método de revisión del estado del arte como lo es la revisión sistemática.

El objetivo es conocer qué indagaciones relacionadas al tema de este trabajo de fin de carrera se han venido realizando en los últimos años, para lo cual se hace uso de la pregunta de investigación la cual nos ayudará en la revisión; de este modo se conocerá el estado actual en el que encuentran y se podrá responder a la interrogante anteriormente mencionada.

El método usado es la revisión sistemática de la literatura el cual tiene como finalidad identificar, evaluar e interpretar toda investigación relevante a una pregunta de investigación particular, o tema de un área, o fenómeno de interés. Estudios individuales que contribuyen a la revisión sistemática son llamados estudios primarios; una revisión sistemática es una forma de estudio secundario [Kitchenham 2004]. Esta revisión se llevó a cabo en diversas bases de datos como: Scopus, *ACM Digital Library* y *IEEE Explore*.

6.1 Formulación de la pregunta

La pregunta que direccionó el trabajo fue: ¿De qué maneras se ha venido realizando la recuperación de videos en el ámbito de la Web Semántica en los últimos años? , para lo cual se usaron las siguientes palabras clave: “*video retrieval*”, “*information retrieval*”, “*semantic Web*” y videos.

6.2 Selección de las fuentes

A partir de la combinación de los términos, que se mencionaron anteriormente y usando los operadores lógicos, se llevó a cabo la revisión con las sentencias que se pueden contemplar en la Tabla 3.

Tabla 3: Cadenas generales de búsqueda avanzada para Scopus, ACM Digital Library y IEEE

	Bases de Datos	Cadenas generales básicas de búsqueda avanzada
1	Scopus	((("semantic web") AND "information retrieval") AND videos)
2		((("semantic web") AND "video retrieval")
3	IEEE Explore	"semantic web" AND "information retrieval" AND videos
4		"semantic web" AND "video retrieval"
5	ACM Digital Library	"information retrieval" and "semantic web" and videos and sports and "semantic annotation"

Para el uso de las cadenas de búsqueda se tomó en cuenta la aparición de las palabras clave dentro del resumen, título y palabras claves para el caso de Scopus y ACM; en cambio para el caso de IEEE se usó la opción solo metadatos.

Se limitó la búsqueda a documentos publicados desde el 2008 hasta la fecha, ya que al querer encontrar documentos relevantes será necesario limitar la búsqueda a los últimos años. De esta forma se conocerá como se encuentran las investigaciones en la actualidad. Esta limitación se usó como criterio de exclusión, el no tomar en cuenta los documentos previos a esa fecha. En la Tabla 4 se muestra se muestra el resumen de los criterios usados para las búsquedas; además de la cantidad de documentos encontrados.

Tabla 4: Resumen de Resultados devueltos por Bases de Datos

Cadena	Bases de Datos	Criterio de Inclusión	Criterio de Exclusión	Número de documentos encontrados
1	Scopus	Sí	Sí	31
2		Sí	Sí	14
3	IEEE Explore	Sí	Sí	10
4		Sí	Sí	17
5	ACM Digital Library	Sí	Sí	62

6.3 Tablas Resumen del estado del arte

6.3.1 Investigaciones

En la Tabla 5 se puede contemplar el resumen de las investigaciones con los diferentes temas que son abordados en estas.

Tabla 5: Resumen de Investigaciones

	Anotaciones	¿Usa Ontologías?	Algoritmo	Indexación y Recuperación de Información	Se diferencia en
[Sebastine 2009]	Sí	Sí (TBOX)	Basado en el <i>Google Normalized Distance</i>	Sí	Se anota el video en cada capa de este, según las características que presente. Uso de TBOX.
[Abel 2008]	Sí	Sí	<i>FolkRank</i> basado en el <i>PageRank</i>	No	Generador de RDF, además uso de algoritmo para recomendar etiquetas. Sistema Social Bookmarking.
[Bertini 2009]	Sí	Sí	No	No	Uso de ontologías dinámicas para la anotación.
[Kannan 2010]	Sí	Sí	Tree-embedding	Sí	Metadatos MPEG-7 y uso de Tree-embedding

6.3.2 Productos

En la Tabla 6 se puede observar el resumen de los productos con sus respectivas tecnologías y una breve descripción de estas.

Tabla 6: Resumen de productos

	Tecnología	Realiza lo siguiente
Blinkx	CORE	Procesamiento, análisis y etiquetado
Faviki	DBpedia	Extraer información estructurada desde Wikipedia
Flipboard	Ellerdale Project	Análisis semántico de flujos de datos en tiempo real
Wolfram Alpha	Motor de conocimiento computacional	Curación de datos, segmentación, sistemas de computación algorítmica, procesamiento lingüístico y el sistema de presentación automatizado

6.4 Investigaciones acerca del tema

En este punto se tomarán en cuenta las investigaciones, que según la revisión hecha se han tomado como relevantes, que se han dado en los últimos años acerca del tema.

6.4.1 Web Semántica para el contenido basado en recuperación de videos

En esta investigación, se denota el uso de la recuperación de videos según el contenido, ya que actualmente se usa la recuperación a través de etiquetas que se colocan manualmente en los videos [Sebastine 2009]. RDF es una tecnología para el desarrollo de la Web Semántica, el cual codifica el significado en conjunto de triplas; siendo las triplas como sujetos, verbos y objetos de una sentencia primaria [Berners-Lee, Hendler et al. 2001]. TBOX es un acrónimo usado en la Comunidad de la Web Semántica para representar Descripción Estructural u Ontología. Para realizar la recuperación de videos según el contenido se tiene el siguiente diseño; en primer lugar, el video se segmenta, en otras palabras, por cada capa del video se extraen sus características; luego, se acumulan todas estas para convertirlas a RDF; después, con ayuda de la ontología TBOX, la cual se acomoda a múltiples estándares, se hace uso de los descriptores de texto y los descriptores de multimedia para describir el tipo de cada característica del video; paso siguiente, con ayuda del sistema de archivos de *Google* se almacenan los archivos RDF; para luego indexar todos los recursos presentes en los archivos RDF, teniendo cada característica del video un índice diferente para facilitar la búsqueda por característica; finalmente con ayuda de un nuevo algoritmo, basado el algoritmo *Google Normalized Distance*, se ordenan los videos de acuerdo a su importancia [Sebastine 2009].

6.4.2 Explotación de un contexto adicional para recomendaciones de etiquetado basado en grafos en sistemas de Folksonomía

En esta investigación se habla de *GroupMe!* el cual es una nueva especie de sistema de recursos compartidos, que extiende la idea de un sistema social de *Bookmarking*,

con la habilidad de crear grupos de recursos multimedia. Como característica importante de *GroupMe!*, el contenido de los grupos es accesible por máquinas como agentes de Web Semántica, ya que es un generador de RDF; además de extraer los datos RDF de los recursos. Otras características, son el agrupamiento y el enriquecimiento de los recursos con semánticas entendibles por las máquinas como RDF generados en *GroupMe!* lo que es hecho de acuerdo a los principios de *Linked Data*.

El etiquetado en *GroupMe!* es realizado manualmente por los usuarios; además que se tiene un sistema de recomendación de etiquetas para los recursos, que ayuda cuando no se tiene ninguna etiqueta en el recurso; para luego ver si estas recomendaciones son relevantes o no; un ejemplo de esto, siendo el recurso un video, es lo siguiente: Usando el algoritmo *FolkRank* basado en el *PageRank* se ordenan los recursos según relevancia dentro del grupo; además que ordena los grupos por relevancia también [Abel 2008].

6.4.3 Ontologías dinámicas enriquecidas gráficamente para bibliotecas digitales

Con el crecimiento de la cantidad de archivos multimedia y el uso de etiquetado basado en el teclado se está mermando la eficacia; por ello, la necesidad de sistemas de anotación y recuperación para incrementar su precisión; para lo cual, las tecnologías de la Web Semántica prometen una gestión de los metadatos asociados con los archivos de una manera más fácil.

Para esto las ontologías juegan un rol fundamental anotando el contenido de las bibliotecas de videos digitales, porque permiten asociarlos con sus conceptos y datos visuales; sin embargo, enlazar los conceptos de las ontologías a los datos visuales posee muchos problemas que aún se está lejos de resolver. Uno de los problemas principales es como obtener una expresión completa del contenido de información de los datos visuales; otro problema clave es el hecho de que las manifestaciones visuales de los eventos y los objetos pueden cambiar con el tiempo.

Para esto, el modelo incluye no solo conceptos lingüísticos, sino también prototipos visuales para considerar todas las modalidades en las cuales los datos visuales se pueden manifestar. Son obtenidos al agrupar las instancias de datos visuales que son observados, de acuerdo a características perceptuales distinguidas. El modelo maneja modificaciones temporales de datos visuales a través de un mecanismo de agrupamiento que puede reagrupar las instancias ya observadas y redefinir los

prototipos visuales. Se usa el lenguaje Web ontología para modelar tanto los conceptos como los prototipos visuales, y las *Semantic Web Rule Language* (SWRL) para manejar los resultados de la clasificación y derivar nuevas anotaciones semánticas [Bertini 2009].

6.4.4 DanVideo: un sistema MPEG-7 de autoría y recuperación de videos de danza

DanVideo es un sistema que se encarga de la autoría y acceso a archivos de danza; por ello, DanVideo permite realizar métodos de anotación y autoría; además de proveer herramientas de recuperación para coreógrafos, bailarines y estudiantes. DanVideo ofrece una herramienta de autoría semiautomática basada en MPEG-7 que toma las anotaciones realizadas en videos de danza por bailarines expertos y produce metadatos MPEG-7; además tiene también un motor de búsqueda que toma las consultas hechas por los usuarios y devuelve la información tomada de los metadatos organizados usando la técnica *tree-embedding* y basada en características espaciales, temporales y espacio-temporales de los bailarines. El motor de búsqueda también posee una ontología específica para el dominio para procesar consultas basadas en conocimiento.

Las consultas de videos de danza y anotaciones semánticas de DanVideo se han evaluado en términos de precisión y *recall* con consultas que envuelven relaciones de contención, temporales, espaciales y espaciotemporales. Las siguientes son las tablas de resultados con las 4 tipos de consultas [Kannan 2010]. En la Figura 5 se puede contemplar la consulta Contención donde se hicieron las consultas CQ22 y CQ02 que fueron “Show all romantic dance steps” y “Show the dance steps of left eye performed by Lisa” respectivamente con los siguientes resultados.

Consulta Contención	Recall	Precision
CQ22	100	66.66
CQ02	100	50

Figura 5: Resultados de las Consultas Contención [Kannan 2010].

En la Figura 6 se puede observar la consulta temporal donde se hicieron las consultas TQ08 y TQ09 que fueron “Show all dance steps where right hand meets left hand of Anitha” y “Show Anitha rotating eyes during Raghav lifting up shoulders” respectivamente con los siguientes resultados.

Consulta Temporal	Recall	Precision
TQ08	100	100
TQ09	100	100

Figura 6: Resultados de las Consultas Temporales [Kannan 2010].

En la Figura 7 se puede advertir la consulta espacial en donde se hicieron las consultas SQ01 y SQ03 que fueron “Show all dance steps where Anitha is to the left of Raghav” y “Show all dance steps where Anitha is to the right of Raghav” respectivamente con los siguientes resultados.

Consulta Espacial	Recall	Precision
SQ01	33.33	100
SQ03	100	100

Figura 7: Resultados de las Consultas Espaciales [Kannan 2010].

En la Figura 8 se puede contemplar la consulta espacio-temporal en donde se hicieron las consultas STQ1 y STQ2 que fueron “Show all steps where Anitha approaches Raghav while Peter is to the left of Anitha” y “Show all steps where Anitha meets Raghav while Peter is to the left of Anitha, rotating his hip” respectivamente con los siguientes resultados.

Consulta Espaciotemporal	Recall	Precision
STQ1	100	100
STQ2	50	100

Figura 8: Resultados de las Consultas Espaciotemporales [Kannan 2010].

6.5 Productos similares al tema específico

En este punto se darán a conocer productos que están relacionados con los temas; además estos productos están en funcionamiento en la actualidad.

6.5.1 Blinkx

Blinkx es un buscador de videos, el cual usa *CORE* el cual está basado en tecnología que fue desarrollada durante 12 años en la Universidad Cambridge y está protegida por 111 patentes. *CORE* automatiza tareas manuales tradicionalmente costosas y que consumen tiempo incluyendo procesamiento, análisis y etiquetado, y expone todo el potencial de contenido de audio y video. Para lograr esto, *CORE* de Blinkx destila el video en bloques de construcción básicos y luego procesa cada componente en el método más eficaz para cada medio; por consiguiente, el audio lo procesa con tecnología de reconocimiento de voz, las imágenes con tecnología de análisis visual, y los *metatags* son indexados y extraídos por arañas visuales distribuidas. Una vez que ha recolectado los datos más completos, en otras palabras, los datos que minuciosamente ha encontrado que son los más precisos en el video, el motor *CORE* los transforma en información útil mediante el reconocimiento de ideas y temas dentro del video [Blinkx Tech].

6.5.2 Faviki

Faviki es una herramienta de *bookmarking* social que usa los conceptos de Wikipedia como etiquetas; de esta manera permite mantener al usuario sus propias etiquetas conectándose a conceptos universales desde la más grande colección de conocimiento del mundo. Al usar etiquetas comunes con conceptos bien definidos desde Wikipedia, te permite saber sobre qué es exactamente una página Web, permitiéndole a la computadora entender al usuario [Faviki]. La tecnología que usa Faviki es DBPedia la cual es un esfuerzo de una comunidad de varias fuentes por extraer información estructurada desde Wikipedia y hacer que esta información se encuentre disponible en la Web, permitiendo hacer consultas sofisticadas a Wikipedia y enlazando los diferentes conjuntos de datos en la Web a los datos de Wikipedia. La base de conocimiento DBpedia solo en su versión inglesa describe 4 millones de cosas, de las cuales 3.22 millones están clasificadas en una ontología consistente, incluyendo 832000 personas, 639000 lugares, 372000 trabajos creativos (álbumes de música, filmes, videojuegos, entre otros), 209000 organizaciones, 226000 especies y 5600 enfermedades. Cabe resaltar que DBpedia versiones localizadas en 119 lenguajes que conjuntamente describen 24.9 millones de conceptos [DBpedia].

6.5.3 Flipboard

El 21 de julio del año 2010 Flipboard adquirió Ellerdale la cual en ese tiempo era una compañía de inteligencia Web en tiempo real. Ellerdale fue fundado en el año 2008, y ha desarrollado tecnología de inteligencia Web que aplica análisis semántico de flujos de datos en tiempo real para extraer información relevante y útil. Con la tecnología de Ellerdale las versiones de Flipboard son capaces de categorizar y ofrecer contenidos altamente relevantes y de último minuto de temas de tendencia en toda la variedad de redes sociales. De esta manera el lector siempre encontrara los contenidos más importantes y de su interés personales [Flipboard 2010].

6.5.4 Wolfram alpha

Wolfram alpha es un motor de conocimiento computacional, el cual genera respuestas haciendo computaciones desde su base interna de conocimiento. La tecnología que usa se divide en muchas partes como son la curación de datos, segmentación, sistemas de computación algorítmica, procesamiento lingüístico y el sistema de presentación automatizado. Aunque no use directamente Web Semántica, lo hace indirectamente con el uso de ontologías y semánticas extensas construidas internamente [WA FAQ].

6.6 Conclusiones sobre el estado del arte

En conclusión, dadas las investigaciones y productos mencionados anteriormente se tiene una idea más clara de cómo poder afrontar la problemática de este proyecto de fin de carrera. De esta forma, analizando la forma en la que plantearon solucionar las diferentes problemáticas en diferentes contextos podemos extraer los pasos o las acciones que realizaron que se adecuen de mejor manera a nuestra problemática.

Por esta razón es importante haber realizado esta revisión sistemática, ya que nos da una idea de qué forma se podría afrontar el problema a partir de los resultados que obtuvieron las investigaciones.



Capítulo 3

7. Resultados Esperados del Objetivo Específico 1

7.1 Ontología que presente el conocimiento en el dominio de las Artes Marciales Mixtas

Para la realización de este resultado esperado se hizo uso de la Metodología para el Desarrollo de Ontologías 101, la cual nos describe 7 pasos, por ello se hará mención de cada uno de estos pasos y el planteamiento que se hizo para realizar la ontología. Para el desarrollo de la ontología se hizo uso de la herramienta Protégé.

7.1.1 Paso 1: Determinar el dominio y alcance de la Ontología

Se recomienda que para iniciar el Desarrollo de una Ontología en primer lugar se haga la especificación del dominio y el alcance que va a tener esta. Para ello se recomienda responder a preguntas básicas como [Noy and McGuinness 2001]:

- ¿Cuál es el dominio que la ontología podría cubrir? [Noy and McGuinness 2001]

El dominio que se cubrirá son las Artes Marciales Mixtas limitándolo a la empresa UFC (Ultimate Fighting Championship).

- ¿Para qué se va a usar la Ontología? [Noy and McGuinness 2001]

La Ontología se usará para la recuperación semántica de videos en el dominio de las Artes Marciales Mixtas. Específicamente se recuperarán peleas.

- ¿Para qué tipo de preguntas la Ontología debe proveer respuestas? [Noy and McGuinness 2001]

La ontología debe poder responder a preguntas que se refieran a temas pertenecientes al dominio, por lo cual la Ontología debe contener conceptos relacionados a este y relacionarlos con los encuentros que tengan relación con lo preguntado.

- ¿Quién podría usar y mantener la Ontología? [Noy and McGuinness 2001]

El uso y mantenimiento de la ontología será hecho por el autor de esta.

Una vez respondidas estas preguntas tendremos una idea más específica del alcance que tendrá nuestra ontología.

7.1.2 Paso 2: Considerar el reúso de Ontologías existentes

Vale la pena considerar lo que otra persona ha hecho y comprobar si podemos perfeccionar o ampliar las fuentes existentes para nuestro dominio y tarea particular [Noy and McGuinness 2001]. Para realizar este paso se buscó en páginas Web ontologías que tratasen del dominio; las paginas usadas para realizar esta búsqueda fueron Swoogle, Ontolingua y DAML las cuales son bibliotecas de ontologías. Desafortunadamente no se encontró ninguna ontología que tratase acerca de las Artes Marciales Mixtas por lo cual no se podrá hacer una reutilización y se tendrá que realizar una Ontología desde cero.

7.1.3 Paso 3: Enumerar términos importantes en la Ontología

Es útil hacer una lista de todos los términos acerca de los que nos gustaría declarar o explicar al usuario. ¿Cuáles son los términos de los que nos gustaría hablar? ¿Qué propiedades tienen estos términos? ¿Que nos gustaría decir sobre estos términos? [Noy and McGuinness 2001]. Para este paso se hizo la lista de los términos más representativos del dominio entre los cuales estaban los siguientes:

- Event
- Fight
- Fighter
- Finalización
- Referee
- Title
- Weight Class.
- Main Card
- Prelims
- Title Fight
- Main Event
- Secondary Event
- Decision
- Disqualification
- Knockout
- Submission
- ChokeHold
- JointLock
- Technical Knockout
- Round

7.1.4 Paso 4: Definir las clases y Jerarquía de Clases

Para el desarrollo de este paso se hizo uso del proceso de desarrollo de arriba hacia abajo el cual comienza desde los conceptos más generales en el dominio para la posterior especialización de los conceptos [Noy and McGuiness 2001]. En la Figura 9 se muestra la Jerarquía de Clases que se realizó teniendo tanto conceptos generales como sub-clases de estos que serían la especialización de estos conceptos.

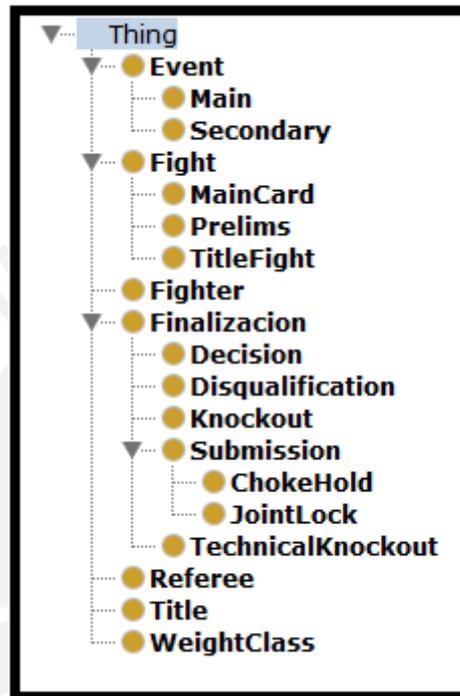


Figura 9: Jerarquía de Clases. Autoría Propia.

7.1.5 Paso 5 y 6: Definir las propiedades de las Clases-Ranuras y Definir las facetas de las Ranuras

Los Clases por sí solas no proveen de información suficiente para responder a las preguntas que se mencionaron que se responderían en el Paso 1, por ello se deberá describir la estructura interna de los conceptos. Esta descripción servirá para definir las propiedades las cuales se vuelven ranuras adjuntas a las clases [Noy and McGuiness 2001]. Las Ranuras tienen diferentes facetas describiendo valores que pueden ser cadenas de caracteres o números o instancias de otras clases; dicho de otra manera, la llamada cardinalidad, además las ranuras pueden contener un Dominio y un Rango [Noy and McGuiness 2001].

Para estos 2 pasos se describieron las Clases y se colocó la cardinalidad necesaria a las propiedades que necesitaban de una. Por ejemplo en el caso de un peleador campeón se hizo uso de la propiedad de objeto “*HasTitle*” la cual posee como Dominio una peleador (*Fighter*) y como Rango un Título (*Title*). Para el caso de si el peleador

contaba con un apodo (*Nickname*) se hizo uso de la propiedad de datos “*Nickname*” la cual tenía de cardinalidad *String*. De esta manera se describieron todas las clases usando tanto propiedades de objeto como propiedades de datos, además de definir las facetas de estas.

7.1.6 Paso 7: Crear Instancias

El último paso es crear una instancia individual de clases en la jerarquía. Definiendo una instancia individual de una clase requiere (1) escoger una clase, (2) crear una instancia individual de esta clase, y (3) completar los valores ranura. Para este paso por ejemplo en la clase *Fighter* se creó la instancia individual Jon Jones la cual poseía las propiedades “*HasTitle*”, “*Nickname*”, “*Weight*”, “*Height*”, entre otras las cuales se llenaron con información que se encuentra disponible en la página Web oficial de la UFC. Finalmente luego de haber creado todas las instancias se obtiene la ontología la cual se puede observar en la Figura 10.

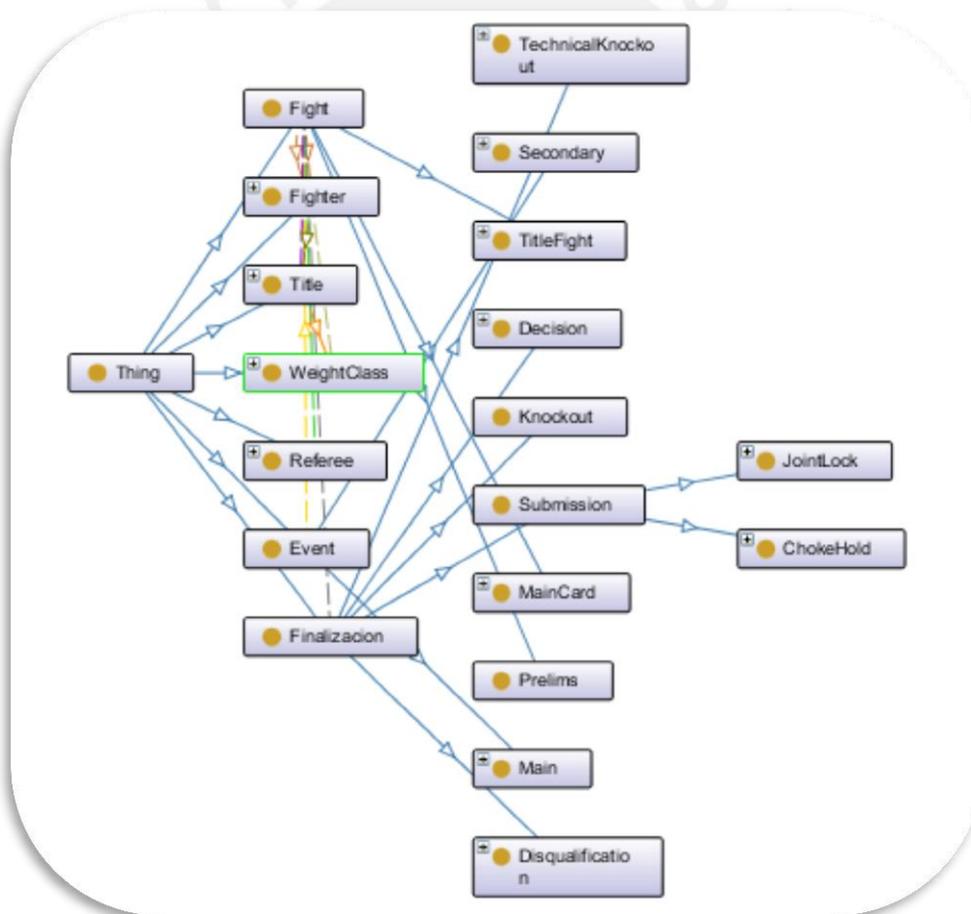


Figura 10: Ontología del dominio de las MMA. Autoría Propia.

7.2 Estructura de anotación semántica de videos.

Para la estructura de anotación semántica tal como se observa en la Figura 11 se planteó el uso de una Base de Datos de Conocimiento en la cual se encontrarían almacenados los URLs de los videos relacionados a URLs de las peleas a las cuales

están relacionadas. El uso de los URIs en la Base de Datos se da para la interacción con la Ontología dado que los URIs son la representación de un individuo de esta. Esta estructura dará paso a que el usuario al realizar una búsqueda en la herramienta, esta se realice primeramente en la ontología la cual por este motivo se recorrerá y se obtendrán los URIs de los encuentros que estén relacionados a la consulta. Finalmente, teniendo ya los URIs se hará un “match” con los videos que se encuentran en la Base de Datos de Conocimiento para obtener los URLs de los videos que contengan información deseada. La base de datos contendrá información que nos permitirá reconocer el contenido del video.

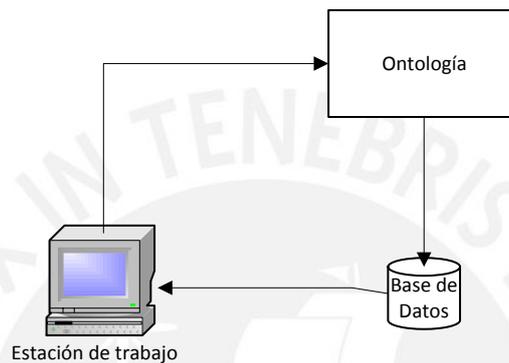


Figura 11: Planteo de funcionamiento de la Herramienta. Autoría Propia.

7.3 Base de datos de videos relacionados al dominio de las Artes Marciales Mixtas.

La Base de Datos fue realizada con la herramienta MySQL *Workbench Community Edition*. Como se observa en la Figura 12 la base de datos posee tan solo una tabla llamada *Fight*, la cual representa un encuentro de Artes Marciales Mixtas. En esta tabla se tiene como llave primaria el URL del video del encuentro, además posee como atributos el URI el cual es la representación de la pelea que posee como contenido el video en la ontología, el nombre del video el cual es el que posee en la página Web en donde está alojado. Los siguientes atributos serán usados como “flags” para filtrar el contenido que se quiere mostrar es así que el *flag* Finalización nos dirá si el contenido del video muestra o no como terminó el encuentro, el *flag* de *Countdown* indicará si el contenido del video es la cuenta regresiva previa a la pelea, el *flag* *Highlight* mostrará si el video contiene el resumen del encuentro, los *flags* Round 1 al 5 nos advertirán si estos están presentes en el video, el *flag* *WeighIn* señalará si el video contiene el pesaje previo a la pelea, el *flag* *Entrance* exhibirá si el video contiene la entrada de los peleadores, y por último el *flag* *FullMatch* indicará si el video posee el encuentro completo. En el Anexo 1 se puede observar el Diccionario de Datos de la tabla *Fight* para una mejor comprensión.

7.4 Herramienta que permite añadir metadatos a los videos usando la ontología y la estructura de anotación creada anteriormente.

La herramienta diseñada para añadir metadatos a los videos usando la ontología y la estructura de anotación creada anteriormente, se realizó en Java en un archivo *main*. Esta herramienta posee como entrada un archivo de texto con los individuos que se desean crear en la base de datos tal como se observa en la Figura 13 y como salida la base de datos con los videos anotados. Es por ello que esta herramienta posee el acceso a la base de datos para poder realizar la creación de los individuos con la información que se le ha otorgado por medio del archivo de texto.



Figura 12: Modelo de Base de Datos. Autoría Propia.



```

input.txt: Bloc de notas
Archivo Edición Formato Ver Ayuda
www.youtube.com/v/bkKEdtPisz8,http://www.semanticweb.org/james/ontologies/2014/6/mma#MainEventUFC175,
UFC 175: Fight Motion,1,0,1,1,1,1,1,1,1,0,0,0
www.youtube.com/v/Tt4Ze8l7Ssg,http://www.semanticweb.org/james/ontologies/2014/6/mma#MainEventUFC175,
UFC 175 Chris Weidman vs Lyoto Machida weigh-in + Stare Down(HD /Unedited),0,0,0,0,0,0,0,0,1,0,0
www.youtube.com/v/5j6td0q8Bu4,http://www.semanticweb.org/james/ontologies/2014/6/mma#MainEventUFC175,
UFC 175 Chris Weidman vs Lyoto Machida walkout,0,0,0,0,0,0,0,0,0,1,0
www.youtube.com/v/6biwfiLXGHU,http://www.semanticweb.org/james/ontologies/2014/6/mma#MainEventUFC175,
Countdown to UFC 175 - Episode complet en version française,0,1,0,0,0,0,0,0,0,0,0
  
```

Figura 13: Archivo de texto que sirve como input a la herramienta. Autoría Propia.

Capítulo 4

8. Resultados Esperados del Objetivo Específico 2

8.1 Herramienta que permite la indexación de los videos usando como representación de los mismos metadatos

Para el desarrollo de este resultado esperado se usó la herramienta Lucene la cual permitió la implementación del indexador. Para esto, también se usó en conjunto la herramienta Jena la cual nos permitió sacar los datos de la Ontología para ser indexados. El uso de estas 2 herramientas se hizo en sus versiones para el lenguaje Java. Para cumplir con la funcionalidad requerida primero se hizo uso de Jena la cual permitió recorrer la ontología y obtener los datos de cada individuo presente en ella lo cual se hizo a través de las propiedades que poseían. Teniendo los datos de los individuos se pasó a llenar documentos con estos, para la creación del indexador con el uso de Lucene. Por último estos documentos se colocan en un *dataset* el cual estará corriendo en RAM cada vez que se ejecute la herramienta.

En la Figura 14 se puede observar cómo se realizó el llenado e indexación de los documentos con la siguiente función la cual tuvo como entradas el indexador y la lista de Datos de un Individuo para su consiguiente indexación.

```

public void indexDocs(IndexWriter writer, List<String> listInd)
throws IOException {
    try {
        //Creación de un documento
        Document doc = new Document();
        String aux;
        //Cadena que contendrá todo el contenido de los datos del individuo
        String content = "";
        //Como los datos entran con un formato Tipo:Dato se realiza el
        //proceso de separar las 2 cadenas
        int pos;
        for(int i=0;i<listInd.size();i++){
            aux = listInd.get(i);
            pos = aux.indexOf(":");
            //Se agrega el dato a la cadena contenido
            content += aux.substring(pos+1, aux.length())+ " ";
            //Se agrega el tipo con su dato en el documento
            doc.add(new Field(aux.substring(0, pos),aux.substring(pos+1,
                aux.length()) , TextField.TYPE_STORED));
        }
        //Se agrega toda la cadena contenido al documento
        doc.add(new Field("contents",content , TextField.TYPE_STORED));
        //Finalmente se agrega el documento al indexador
        writer.addDocument(doc);
    }
}

```

Figura 14: Codificación realizada para la Indexación. Autoría Propia.

8.2 Herramienta que permite recuperar los videos usando búsqueda textual

La herramienta para la recuperación de videos usando búsqueda textual se implementó usando la herramienta Lucene. La función que cumple esta herramienta es que al recibir un *input* debe verificar que este se encuentre en los datos que contiene nuestro indexador para devolvernos como respuesta los individuos que tengan relación con este. De esta manera se podrá encontrar los individuos con el uso de consultas textuales para usarlos en la consiguiente búsqueda semántica. En la Figura 15 se puede contemplar cómo se pasa a buscar en el directorio los documentos que en este caso con los individuos que estén relacionados con la cadena de búsqueda "line".

```

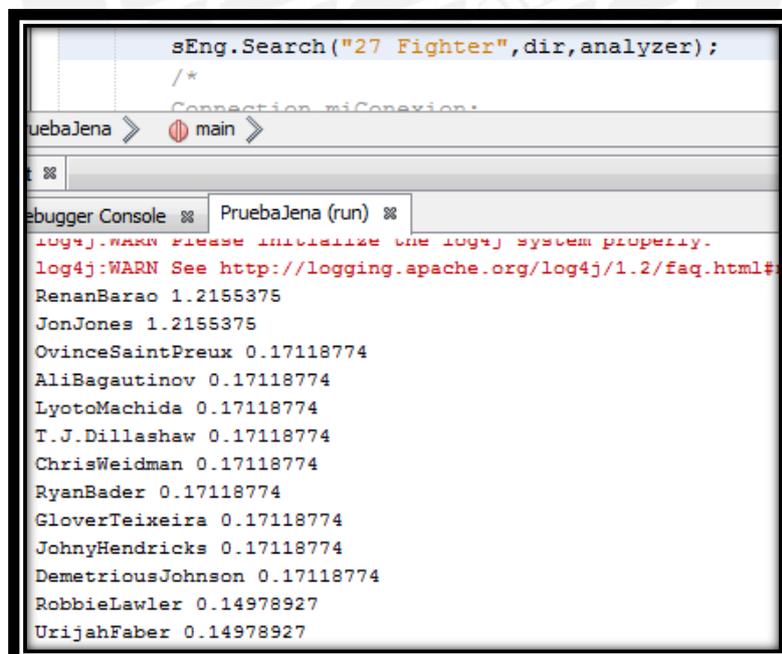
public List<String> Search(String line, Directory dir, Analyzer analyzer)
    throws IOException, ParseException {
    String field = "contents";
    List<String> result = new ArrayList<String>();
    DirectoryReader ireader = DirectoryReader.open(dir);
    IndexSearcher isearcher = new IndexSearcher(ireader);
    QueryParser parser = new QueryParser(Version.LUCENE_4_10_0, field, analyzer);
    Query query = parser.parse(line);
    ScoreDoc[] hits = isearcher.search(query, null, 1000).scoreDocs;
    for (int i = 0; i < hits.length; i++) {
        Document hitDoc = isearcher.doc(hits[i].doc);
        result.add(hitDoc.get("Name"));
        System.out.print(hitDoc.get("Name")+" "+ hitDoc.get("Class") + " ");
        System.out.print(hits[i].score);
        System.out.println();
    }
    ireader.close();
    dir.close();

    return result;
}

```

Figura 15: Codificación realizada para la Búsqueda. Autoría Propia.

Finalmente en la Figura 16 se hace la búsqueda de la cadena “27 Fighter” para lo cual el motor de búsqueda nos devuelve los nombres de los individuos con un score al costado. Este score nos permite conocer cuánto tiene que ver la cadena de búsqueda en relación con el individuo, por ello mientras mayor sea el número del score mayor es la relevancia de esa respuesta.



```

sEng.Search("27 Fighter", dir, analyzer);
/*
Connection miConexion:
uebalena > main >
Debugger Console || PruebaJena (run) ||
log4j:WARN Please initialize the log4j system properly.
log4j:WARN See http://logging.apache.org/log4j/1.2/faq.html#
RenanBarao 1.2155375
JonJones 1.2155375
OvinceSaintPreux 0.17118774
AliBagautinov 0.17118774
LyotoMachida 0.17118774
T.J.Dillashaw 0.17118774
ChrisWeidman 0.17118774
RyanBader 0.17118774
GloverTeixeira 0.17118774
JohnnyHendricks 0.17118774
DemetriousJohnson 0.17118774
RobbieLawler 0.14978927
UrijahFaber 0.14978927

```

Figura 16: Respuesta de una Búsqueda Textual. Autoría Propia.

8.3 Herramienta que permite recuperar los videos usando búsqueda semántica

Esta herramienta se realizó con el uso de Jena el cual nos permitió, con el resultado anterior, recorrer la ontología para encontrar las peleas relacionadas a los individuos o individuo el cual se buscó. Este recorrido de la ontología a través de las relaciones se hizo aprovechando las propiedades que posee cada individuo. Es así como finalmente luego de haber obtenido las peleas en las cuales participaron los individuos buscados estas se devuelven mediante el uso de sus URIs. Estas URIs son usadas para luego hacer el *match* con las que se encuentran en la base de datos y obtener los URLs de los videos de las peleas.

En la Figura 17 se puede observar una función que a partir de un individuo de cualquier clase puede llegar a obtener los encuentros en los que ha participado.

```

public List<String> FightSearchOthers(String individuo){
    ExtendedIterator<? extends OntResource> fights = getIndividuals("Fight");
    List<String> result = new ArrayList<String>();
    while(fights.hasNext()){
        OntResource ind = (OntResource)fights.next();
        StmtIterator it2 = ind.listProperties();
        while (it2.hasNext()) {
            Statement s = (Statement) it2.next();
            Property p = s.getPredicate();
            OntProperty oP = ontModel.getOntProperty(namespace +
                p.getLocalName());
            if(oP!=null && oP.isObjectProperty()){
                Resource resource = s.getObject().asResource();
                if(resource.getLocalName().equals(individuo)){
                    if(!result.contains(ind.getLocalName())){
                        result.add(namespace + ind.getLocalName());
                    }
                }
            }
        }
    }
    return result;
}

```

Figura 17: Función que permite acceder desde un individuo a una lucha. Autoría Propia.

Finalmente como se ve en la Figura 18 se obtuvo la herramienta la cual por medio de una búsqueda semántica permitirá obtener videos de Artes Marciales Mixtas.

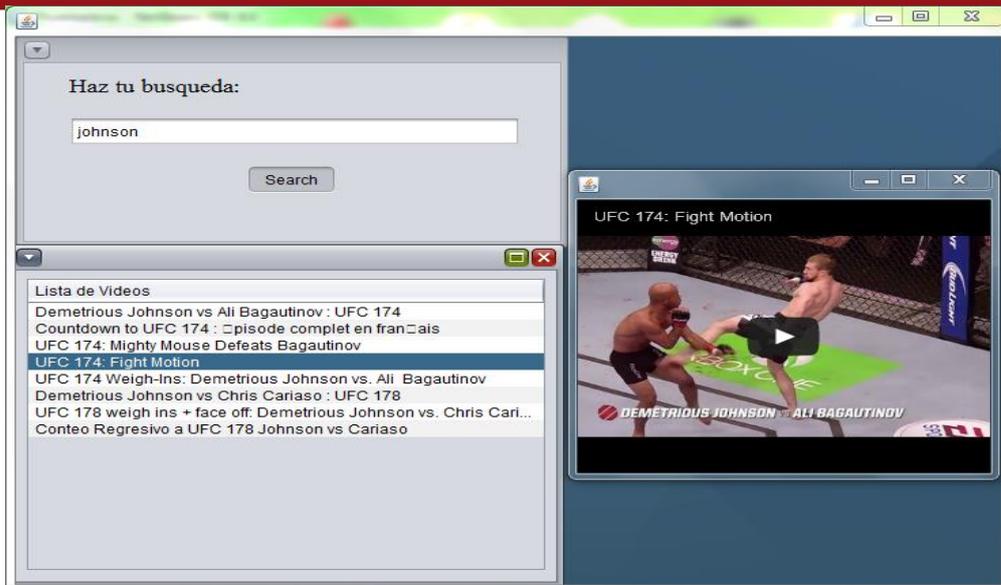


Figura 18: Herramienta que permite recuperar videos. Autoría Propia.

8.4 Precisión y Recall de mecanismo de recuperación de videos

Los resultados acerca de la performance de la herramienta se medirán usando la precisión y el *recall*; para esto se probará la herramienta con diferentes consultas y se verá los resultados en cada una de estas. Para estas pruebas se usaron 50 videos de prueba tanto de Youtube, como de otras páginas similares. Los encuentros se dieron en los eventos UFC 169 al UFC 175, exceptuando el UFC 170. Para mayor detalle de cómo se halló la precisión y el *recall* revisar el Anexo 2.

Tabla 7: Consultas con sus respectivas precisiones y recall

	Consulta	Precisión	Recall
1	Jon Jones	100%	100%
2	UFC On FX 5 Full Match	100%	100%
3	Finalizacion Unanimous Decision	50%	100%
4	Finalizacion Referee Stops	20%	100%
5	Flyweight	100%	100%
6	Middleweight Title	18%	100%
7	UFC 169 Round 1	12%	100%
8	Barao Entrance	100%	100%
9	Herb Dean	100%	100%
10	Main Event UFC 174	10%	100%

Como se observa en la Tabla 7 las búsquedas que poseen baja precisión se dan por la primera búsqueda textual que se realiza, ya que al buscar por ejemplo UFC On FX 5 esta obtendrá los eventos que contengan la palabra UFC, los cuales luego por la búsqueda semántica se obtendrán todas las peleas relacionadas a estos eventos. Por esta razón, nos devuelve todos los videos, pero al encontrarse la cadena “Full Match”, que es un filtro, disminuye la cantidad de videos aumentando así en cierto grado la precisión.



CAPITULO 5

9. Resultados Esperados del Objetivo Específico 3

9.1 Inferencias que permiten realizar búsquedas más específicas

El desarrollo de las inferencias se realizó con la herramienta Jena la cual a partir de las nuevas posibilidades de recuperación fundadas nos permitió obtener los resultados deseados. Una de las posibilidades de recuperación que se plantearon fueron el que debería suceder si el usuario de la herramienta realiza una consulta en la cual encontramos un individuo de clase *Fighter* y la palabra *Knockout*. Es así como se decidió que al desarrollar una consulta de esta forma se deberían encontrar las peleas en las que el luchador haya realizado un *Knockout* a otro. De esta manera se procedió a desarrollar en Jena funciones que nos permitirán realizar búsquedas de este tipo.

Otra posibilidad que se propuso fue el que debería acontecer si en la consulta del usuario encontramos un individuo de clase *Fighter* y la palabra *Submission*. Nuevamente siendo este un caso similar al anterior se decidió que se deberían mostrar los videos en los cuales el luchador haya realizado una *Submission* a su contrincante.

En la Figura 19 se puede observar parte de la codificación que se realizó para que la herramienta pueda realizar estas inferencias, con usar como input el tipo de finalización al que se quería llegar siendo en este caso o un *knockout* o una *submission*.

```

public List<String> FindFighterKnockOrSub(String individuo,String typeFinalizacion){
    List<String> result = new ArrayList<String>();
    List<String> fights= new ArrayList<String>();
    List<String> classes= new ArrayList<String>();
    fights = FightSearchOthers(individuo);
    String valAux = "";
    for(int i=0;i<fights.size();i++){
        OntResource ind = ontModel.getOntResource(fights.get(i));
        Property property = ontModel.getProperty(nameSpace + "WithWinner");
        StmtIterator it = ind.listProperties(property);
        while (it.hasNext()) {
            Statement s = (Statement) it.next();
            Resource resource = s.getObject().asResource();
            valAux = resource.getLocalName();
        }
        if(valAux.equals(individuo)){
            OntResource ind2 = ontModel.getOntResource(fights.get(i));
            Property property2 = ontModel.getProperty(nameSpace + "EndingBy");
            StmtIterator it2 = ind2.listProperties(property2);
            while (it2.hasNext()) {
                Statement s2 = (Statement) it2.next();
                Resource resource2 = s2.getObject().asResource();
                classes = getIndClasses(resource2.getLocalName());
                for(int j=0;j<classes.size();j++){
                    if(classes.get(j).equals(typeFinalizacion)
                        if(!result.contains(fights.get(i)))
                            result.add(fights.get(i));
                }
            }
        }
    }
}
    
```

Figura 19: Codificación realizada para encontrar un peleador que realizó un *knockout* o una *submission*. Autoría Propia.

9.2 Precisión y Recall de mecanismo de recuperación de videos

Dadas las nuevas inferencias que se han desarrollado pasaremos a realizar pruebas a las nuevas consultas que permite desarrollar la herramienta; para ello, nuevamente como realizamos en el resultado esperado 8, hallaremos la precisión y el *recall* lo cual nos ayudará a obtener ratios relacionados a la cantidad de videos relevantes recuperados por la herramienta. Estas pruebas se realizaron con 50 videos y sus resultados se pueden observar en la Tabla 8. Para mayor detalle de cómo se halló la precisión y el recall revisar el Anexo 2.

Tabla 8: Consultas con sus respectivas precisiones y recall

	Consulta	Precisión	Recall
1	Johnson Submission	100%	100%
2	Dillashaw Knockout	100%	100%

CAPITULO 6

10. Conclusiones y Trabajos Futuros

10.1 Conclusiones

Dado el desarrollo en el ámbito de la Web Semántica de una herramienta de recuperación de videos en el dominio de las Artes Marciales Mixtas, se mostrarán las conclusiones obtenidas al haber desarrollado los objetivos específicos que se plantearon previos a implementar la herramienta.

10.1.1 Conclusiones del Objetivo Específico 1

En el objetivo específico 1 inicialmente se hizo el desarrollo de una Ontología que representa el conocimiento del dominio de las Artes Marciales Mixtas. Para el desarrollo de esta ontología se hizo uso de la herramienta Protégé y de la Metodología para Desarrollo de Ontologías 101. Como resultado final al haber seguido los pasos que nos presenta la Metodología se llegó a construir la ontología la cual posee clases, propiedades e individuos que nos permitirán realizar búsquedas a través del conocimiento que contienen. Al desarrollar la ontología se notó como, al ya haber estudiado la metodología y al haber aprendido a usar herramienta Protégé; además de ya conocer el dominio, se hizo un proceso fluido, pero engorroso al tener que crear los individuos y relacionarlos a través de propiedades a datos o a otros individuos.

Al desarrollar la estructura de anotación semántica había que tener una idea clara de cómo se quería que funcionara la herramienta. Teniendo en claro el funcionamiento de la herramienta se procedió a plantear de qué manera con el uso de la Ontología se podría llegar a la finalidad deseada. Finalmente se optó por el desarrollo de una base de datos la cual nos permitiría mapear los videos con los individuos de la Ontología que los representen. De esta manera se pudo acceder de una manera más rápida a los videos al tener los URI de los encuentros.

Como paso siguiente se desarrolló la base de datos relacionada al dominio de las Artes Marciales Mixtas. Esta hizo con ayuda de la herramienta MySQL y se planteó como una tabla la cual contiene el mapeo de las URL con los URI; además de los *flags*. Estos *flags* nos permiten anotar los videos, de esta manera si el video contiene imágenes del Round 1 el *flag* relacionado se encontrará con valor 1. El desarrollo de esta base de datos fue importante ya que simula una tabla hash en donde los URI que nos daba como solución la herramienta nos funcionan como llaves para obtener los URL de los videos.

Para la anotación se tuvo que crear una herramienta que nos permitiera a partir de un documento de texto ingresar datos a nuestra base de datos. Para esto se realizó una función en Java que nos permitiera realizar este cometido. De esta manera se pudo agilizar el proceso de anotación.

10.1.2 Conclusiones del Objetivo Específico 2

Como primer resultado de este objetivo específico se realizó la indexación de los videos usando como representación de los mismos metadatos; se hizo con las herramientas Lucene y Jena las cuales tuvieron una curva de aprendizaje alta al darnos diferentes posibilidades de solución como SPARQL y Query text ,pero finalmente se terminó usando el Ontology API. De esta manera, se realizó el indexado; con lo cual, se tenía la información de los distintos individuos que contiene la ontología en documentos para su posterior búsqueda. Este proceso de indexado, gracias a las librerías que brinda Lucene, se desarrolló rápidamente encontrándose la dificultad en el acceso a los individuos y a sus datos a través de Jena.

Consiguientemente se realizó la búsqueda textual con ayuda de la herramienta Lucene. Esta se realizó sin problemas dado que una vez que se tiene la indexación el proceso se reduce a realizar búsquedas en esta. Es así que a la búsqueda realizada con Lucene se le puede configurar si se requiere de un ranking de los resultados entregados, así como del score que permite conocer cuan parecida fue la cadena de búsqueda a un individuo de la ontología.

Como paso posterior se realizó la búsqueda semántica, para lo cual con ayuda de la herramienta Jena y de los resultados obtenidos de la búsqueda anterior, se obtuvieron los encuentros en los cuales hubo participación del individuo encontrado. Es así que recorriendo la ontología se lograron obtener a través de las relaciones estas peleas. Este proceso fue uno de los más complicados al tener que acceder desde cualquier clase de individuo a un individuo de la clase *Fight*; para lo cual, se desarrollaron diferentes funciones de acuerdo al lugar jerárquico en el que se encontraba el individuo en la ontología.

Finalmente, para conocer resultados de la eficiencia de la herramienta se pasaron a obtener la precisión y *recall* de diferentes consultas que se le podían realizar. Es así como se vio que al tener el *recall* de 100% en todas era claro que se estaban obteniendo todos los videos que eran relevantes para el usuario. Pero en el caso de la precisión en algunas consultas se podían ver valores altos y en otras valores bajos,

esto es dado a la primera búsqueda textual que se realiza donde al tener indexados metadatos, se da que hay palabras que son importantes para varios individuos.

10.1.3 Conclusiones del Objetivo Específico 3

Para la realización del objetivo específico 3 en primer lugar se desarrollaron inferencias que nos permitieran búsquedas más específicas, para esto se tenía que tener en claro que otras formas de buscar se requerían contemplar. Es así que se decidió que al referirse la cadena de búsqueda a un individuo peleador y a un individuo *knockout* o *submission* se buscaría los videos de los encuentros en los que el peleador realizó uno de estos movimientos. Es así como se aumentó el umbral de posibilidades de búsqueda que ya poseía la herramienta.

Finalmente se realizó la medición de la precisión y el *recall* de la herramienta en donde se obtuvieron resultados del 100% en precisión dado la poca cantidad de peleas que han terminado en *knockout* o *submission* recientemente. De esta manera se pudo conocer la eficiencia de la herramienta cuando se usan estas 2 nuevas inferencias.

10.2 Trabajos Futuros

Dado el desarrollo de la herramienta han quedado ciertas cuestiones que se podrían dar en trabajos futuros que podrían permitir una mejor performance de esta. Por ello se pasarán a mencionar los aspectos que podrían ser abordados.

Se pueden realizar mejoras a la ontología con aspectos como poder contener a las diferentes empresas de Artes Marciales Mixtas que existen en la actualidad en diferentes países; así como también colocar más peleadores de diferentes épocas con las fechas en las que estuvieron contratados en las empresas. De esta forma se podría controlar el hecho de que un peleador estuvo en distintas empresas durante toda su carrera como luchador.

Otro aspecto que se podría abordar es el de desarrollar una herramienta que permita la anotación automática de los videos; para de esta forma agilizar este proceso; esto quizás podría implementarse junto a una herramienta que permita el reconocimiento de patrones.

Un aspecto que puede verse complementado con lo anterior es realizar tantas anotaciones como capas o *frames* tenga el video; de esta manera se podrá mejorar la calidad de la anotación al contener hechos que sucedieron en diferentes momentos del video.

Otro trabajo a futuro que se podría realizar es el de hacer la herramienta Web para de esta forma dar más acceso a la herramienta, usando quizás ya no Jena sino JOWL.

El desarrollo de la herramienta también puede ser realizadas con otras API que brinda Jena como lo son Query Text y SPARQL, usando estas 2 herramientas se podrían tener resultados diferentes a los obtenidos, por lo cual servirían para realizar comparaciones entre sus performances.

La velocidad de la entrega de soluciones es otro aspecto que se podría abordar, además de la posible realización de una herramienta que además de recuperar videos pueda recuperar tanto imágenes o documentos, haciéndola de esta manera una herramienta versátil.



Referencias bibliográficas

Papers y Libros

- [Abel 2008] Abel, F., et al. (2008). Exploiting Additional Context for Graph-Based Tag Recommendations in Folksonomy Systems. Web Intelligence and Intelligent Agent Technology, 2008. WI-IAT '08. IEEE/WIC/ACM International Conference on
- [Agosti and Ferro 2007] Agosti, M. and N. Ferro (2007). "A formal model of annotations of digital content." ACM Trans. Inf. Syst. **26**(1): 3.
- [Antoniou 2004] Antoniou, G. (2004). A semantic web primer, the MIT Press.
- [Baeza-Yates and Ribeiro-Neto 1999] Baeza-Yates, R. A. and B. Ribeiro-Neto (1999). Modern Information Retrieval, Addison-Wesley Longman Publishing Co., Inc.
- [Berners-Lee, Hendler et al. 2001] Berners-Lee, T., J. Hendler, et al. (2001). "The semantic web." Scientific American.
- [Bertini 2009] Bertini, M., et al. (2009). "Dynamic Pictorially Enriched Ontologies for Digital Video Libraries." MultiMedia, IEEE **16**(2): 42-51.
- [Bhogal, Macfarlane et al. 2007] Bhogal, J., A. Macfarlane, et al. (2007). "A review of ontology based query expansion." Information Processing & Management **43**(4): 866-886.
- [Cerulo and Canfora 2004] Cerulo, L. and G. Canfora (2004). "A taxonomy of information retrieval models and tools." Journal of Computing and Information Technology **12**(3): 175-194.

- [Corcho 2006] Corcho, O. (2006). "Ontology based document annotation: trends and open research problems." International Journal of Metadata, Semantics and Ontologies 1(1): 47-57
- [Greengrass 2000] Greengrass, E. (2000). Information Retrieval: A Survey.
- [Gruber 1995] Gruber, T. R. (1995). "Toward principles for the design of ontologies used for knowledge sharing?" International Journal of Human-Computer Studies 43(5-6): 907-928.
- [Ingwersen 1992] Ingwersen, P. (1992). Information retrieval interaction, Taylor Graham Publishing.
- [Kannan 2010] Kannan, R., et al. (2010). "DanVideo: an MPEG-7 authoring and retrieval system for dance videos." Multimedia Tools Appl. 46(2-3): 545-572.
- [Kiryakov, Popov et al. 2004] Kiryakov, A., B. Popov, et al. (2004). "Semantic annotation, indexing, and retrieval." Web Semantics: Science, Services and Agents on the World Wide Web 2(1): 49-79.
- [Kitchenham 2004] Kitchenham, B. (2004). "Procedures for performing systematic reviews." Keele, UK, Keele University 33: 2004.
- [Marshall 1997] Marshall, C. C. (1997). Annotation: from paper books to the digital library. Proceedings of the second ACM international conference on Digital libraries. Philadelphia, Pennsylvania, USA, ACM: 131-140.
- [Marshall 1998] Marshall, C. C. (1998). Toward an ecology of hypertext annotation. Proceedings of the ninth

- ACM conference on Hypertext and hypermedia : links, objects, time and space---structure in hypermedia systems: links, objects, time and space---structure in hypermedia systems. Pittsburgh, Pennsylvania, USA, ACM: 40-49.
- [Meena, Kumar et al. 2004] Meena, E., A. Kumar, et al. (2004). An extensible framework for efficient document management using RDF and OWL. Proceedings of the Workshop on NLP and XML (NLPXML-2004): RDF/RDFS and OWL in Language Technology, Association for Computational Linguistics: 51-58.
- [Mitra and Chaudhuri 2000] Mitra, M. and B. B. Chaudhuri (2000). "Information Retrieval from Documents: A Survey." Inf. Retr. **2**(2-3): 141-163.
- [Mitra, Singhal et al. 1998] Mitra, M., A. Singhal, et al. (1998). Improving automatic query expansion. Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval. Melbourne, Australia, ACM: 206-214.
- [O'Hara and Sellen 1997] O'Hara, K. and A. Sellen (1997). A comparison of reading paper and on-line documents. Proceedings of the ACM SIGCHI Conference on Human factors in computing systems. Atlanta, Georgia, USA, ACM: 335-342.
- [O'Leary 1998] O'Leary, D. E. (1998). "Using AI in knowledge management: knowledge bases and ontologies." IEEE Intelligent Systems and their Applications **13**(3): 34-39.
- [Paralic and Kostial 2003] Paralic, J. and I. Kostial (2003). "Ontology-based information retrieval." Information and Intelligent Systems, Croatia: 23-28.

- [Rijsbergen 1979] Rijsbergen, J. v. (1979). Information Retrieval, London: Butterworths.
- [Salton 1981] Salton, G. (1981). "A blueprint for automatic indexing." SIGIR Forum **16**(2): 22-38.
- [Sannomiya, Amagasa et al. 2001] Sannomiya, T., T. Amagasa, et al. (2001). "A framework for sharing personal annotations on web resources using XML." Aust. Comput. Sci. Commun. **23**(6): 40-48.
- [Schreiber 2000] Schreiber, G. (2000). Knowledge engineering and management: the CommonKADS methodology, MIT press.
- [Sebastine 2009] Sebastine, S. C., et al. (2009). Semantic Web for Content Based Video Retrieval. Semantic Computing, 2009. ICSC '09. IEEE International Conference on.
- [Singhal 2001] Singhal, A. (2001). "Modern information retrieval: A brief overview." IEEE Data Eng. Bull. **24**(4): 35-43.
- [Srivastava and Velegrakis 2007] Srivastava, D. and Y. Velegrakis (2007). Intensional associations between data and metadata. Proceedings of the 2007 ACM SIGMOD international conference on Management of data. Beijing, China, ACM: 401-412.
- [Studer, Decker et al. 2000] Studer, R., S. Decker, et al. (2000). "Situation and perspective of knowledge engineering." Knowledge Engineering and Agent Technology.
- [Uren, Cimiano et al. 2006] Uren, V., P. Cimiano, et al. (2006). "Semantic annotation for knowledge management:

Requirements and a survey of the state of the art." Web Semantics: Science, Services and Agents on the World Wide Web 4(1): 14-28.

[Noy and McGuinness 2001] Noy, N. F. and D. L. McGuinness (2001). Ontology development 101: A guide to creating your first ontology, Stanford knowledge systems laboratory technical report KSL-01-05 and Stanford medical informatics technical report SMI-2001-0880.

[Wolfe 2000] Wolfe, J. L. (2000). Effects of annotations on student readers and writers. Proceedings of the fifth ACM conference on Digital libraries. San Antonio, Texas, USA, ACM: 19-26.

Páginas Web

[Blinkx Tech] "Technology"
Disponible en:
<http://www.blinkx.com/corporate/about/technology>

[CommonKADS DEF] "What is CommonKADS"
Disponible en:
<http://www.commonkads.uva.nl/frameset-commonkads.html>

[DBpedia] "DBpedia"
Disponible en:
<http://dbpedia.org/About>

[ESPN 2011] "Silva vs Sonnen"
Disponible en:

http://assets.espn.go.com/photo/2011/1020/mma_u_silva_sonnen_cr_576.jpg

[Faviki]

“Faviki”

Disponible en:

<http://www.faviki.com/pages/welcome/>

[Flipboard 2010]

“Flipboard Acquires Ellerdale to Boost Content Relevancy in New Social Magazine”

21 de julio, 2010, Palo Alto, CA

Disponible en:

<https://flipboard.com/newsroom/releases/flipboard-acquires-ellerdale-to-boost-content-rele/>

[Jena Apache]

“Jena”

Disponible en:

https://jena.apache.org/about_jena/about.html

[Lucene Core]

“Apache Lucene Core”

Disponible en:

<http://lucene.apache.org/core/>

[MySQL COMM]

“MySQL Community Edition”

Disponible en:

<http://www.mysql.com/products/community/>

[OWL W3C]

“OWL”

Disponible en:

<http://www.w3.org/2001/sw/wiki/OWL>

[OWL2 W3C]

“OWL 2”

Disponible en:

<http://www.w3.org/TR/owl2-overview/>

[Protege Stanford]

“Protege”

Disponible en:

<http://protegewiki.stanford.edu/wiki/Protege>

[UFC GANAR]

“Formas de Ganar”

Disponible en:

<http://www.ufcespanol.com/discover/sport/ways-to-win>

[UFC SPORT]

“El Deporte”

Disponible en:

<http://www.ufcespanol.com/discover/sport>

[UFC RULES]

“Reglas Unificadas y Normas”

Disponible en:

<http://www.ufcespanol.com/discover/sport/rules-and-regulations>

[UNIF RULES]

“Reglas Unificadas y Normas de la Asociación de las Comisiones de Boxeo”

Disponible en:

http://www.abcboxing.com/committee_unified_mma_rules.html

[WA FAQ]

“Frequently Asked Questions”

Disponible en:

<https://www.wolframalpha.com/faqs.html>