MECANISMO FORMAL DE CLASIFICACIÓN AUTOMÁTICA DE UN REPOSITORIO DE OBJETOS DE APRENDIZAJE, A PARTIR DE UNA VALORACIÓN INDIVIDUAL DE LOS OBJETOS, CON EL FIN DE MANTENERLOS EN UN CICLO PERMANENTE DE MEJORAMIENTO

Julián Esteban Gutiérrez Posada

Universidad Autónoma de Bucaramanga (UNAB)

Universidad Oberta de Catalunya (UOC)

Bucaramanga, Santander

Febrero 2010

MECANISMO FORMAL DE CLASIFICACIÓN AUTOMÁTICA DE UN REPOSITORIO DE OBJETOS DE APRENDIZAJE, A PARTIR DE UNA VALORACIÓN INDIVIDUAL DE LOS OBJETOS, CON EL FIN DE MANTENERLOS EN UN CICLO PERMANENTE DE MEJORAMIENTO

Julián Esteban Gutiérrez Posada

Tesis de grado para optar el título de:

Magister en E-Learning

Director:

PhD. Miguel Francisco Crespo Alvarado

Docente de la Maestría en E-Learning

Universidad Autónoma de Bucaramanga (UNAB)

Universidad Oberta de Catalunya (UOC)

Bucaramanga, Santander

Febrero 2010

Nota de Aceptación

_____

_____

_____

_____

Presidente del Jurado

_____

Jurado

_____

Jurado

_____

Bucaramanga Febrero de 2010

Dedicatoria

Le dedico esta tesis a

todas aquellas personas que tienen

un espíritu joven y una mente abierta

para pensar que ninguna meta

es inalcanzable por ellos.

Agradecimientos

Quiero expresar mis agradecimientos a:

- Dios quien me dio la paciencia, la dedicación y la fuerza para trabajar duro.
- A mi familia por aceptar y comprender ese gran número de noches de trabajo y especialmente a mi Esposa que tanto AMO.
- A mi director quien me orientó y asesoró incondicionalmente, incluso un semestre antes de inicial formalmente la tesis.
- Al profesor Hernando Hurtado por sus orientaciones e indicaciones durante el desarrollo de la tesis.
- y finalmente a todas aquellas personas que de una u otra forma ayudaron a la realización de esta tesis.

Gracias, muchas gracias a todos y que Dios los bendiga.

ÍNDICE GENERAL

Ficha Resumen **19**

Resumen **21**

1. Introducción **23**

2. Descripción del Problema **25**

3. Objetivos **29**

3.1. Objetivo general 29

3.2. Objetivos específicos 30

4. Estado del arte **31**

5. Marco teórico **37**

5.1. Análisis de Cluster 37

5.1.1. Tipos de datos en el análisis de clases 39

5.1.1.1. Variables basadas en intervalos 40

5.1.1.2. Variables binarias 42

5.1.1.3. Variables nominales 45

5.1.1.4. Variables ordinales 46

5.1.1.5. Variables de tipos mixtos 48
5.1.1.6. Variables tipo vector 50
5.1.2. Métodos de clasificación (Clustering) 51
5.1.2.1. Métodos de particionamiento 51
5.1.2.2. Métodos jerárquicos 52
5.1.2.3. Métodos basados en densidad 52
5.1.2.4. Métodos basados en rejilla 53
5.1.2.5. Métodos basados en modelos 53
5.1.2.6. Métodos de datos de altas dimensiones 53
5.1.2.7. Métodos basados en restricciones 54
5.2. Objetos de aprendizaje 54
5.2.1. Características de un objeto de aprendizaje 55
5.2.2. Metadatos de un objeto de aprendizaje 56
5.2.2.1. Estructura de un metadato 56
5.3. Proceso de medición 57
5.3.1. Modelo de calidad 60
5.3.2. Medida 60
5.3.3. Formas de medir 61
5.3.4. Acción de medir 62

6. Mecanismo formal de clasificación 65
6.1. Espacio de medida 65
6.2. Método de clasificación 68
6.2.1. Condiciones iniciales 68
6.2.2. Condiciones para adicionar un objeto de aprendizaje O 69
6.2.3. Condiciones para retirar un objeto de aprendizaje O 69
6.2.4. Clasificación de los objeto de aprendizaje O 69
6.2.5. Requerimiento sobre el algoritmo 70
6.3. Contexto concreto 71
6.4. Mecanismo de valoración de los indicadores 76

6.5. Ejemplo del método de clasificación 76
6.5.1. Condiciones iniciales 76
6.5.2. Objetos de aprendizaje 77
6.5.3. Repositorio de objetos de aprendizaje 77
6.5.4. Aglomeraciones de los objetos de aprendizaje 78
6.6. Comentario final 80

Conclusiones y Trabajo futuro **81**

Bibliografía **83**

ÍNDICE DE CUADROS

5.1. Matriz de datos 39
5.2. Matriz de disimilaridad 40
5.3. Ejemplo - Variables basados en intervalos 41
5.4. Distacias sobre el ejemplo de variables basadas en intervalos 42
5.5. Variables binarias 43
5.6. Ejemplo de variables binarias 44
5.7. Distacias sobre el ejemplo de variables binarias 44
5.8. Ejemplo de variables nominales 45
5.9. Distancias sobre el ejemplo de variables nominales 46
5.10. Ejemplo de variables ordinales 47
5.11. Enumeración de los valores de las variables ordinales 47
5.12. Transformado la matriz de datos de las variables ordinales 47
5.13. Transformado la matriz de datos de las variables ordinales 48
5.14. Ejemplo de variables de tipos mixtos 49
5.15. Distancias sobre el ejemplo de variables de tipos mixtos 49
5.16. Distacias sobre el ejemplo de variables tipo vector 50

ÍNDICE DE FIGURAS

5.1. Ejemplo - Variables basadas en intervalos 41
5.2. Diagrama de la Ontología de la Medición del Software 58

FICHA RESUMEN

Título:

MECANISMO FORMAL DE CLASIFICACIÓN AUTOMÁTICA DE UN REPOSITORIO DE OBJETOS DE APRENDIZAJE, A PARTIR DE UNA VALORACIÓN INDIVIDUAL DE LOS OBJETOS, CON EL FIN DE MANTENERLOS EN UN CICLO PERMANENTE DE MEJORAMIENTO

Director: PhD. Miguel Francisco Crespo Alvarado

Investigador: Julián Esteban Gutiérrez Posada

Línea de Investigación: E-learning para la educación superior.

Facultad: Educación.

Total de Investigadores: 1

Entidad: Universidad Autónoma de Bucaramanga en convenio con: Universitat Oberta de Catalunya

Lugar de ejecución del proyecto: Universidad del Quindío, Armenia - Quindío.

Ciudad: Armenia.

Duración del proyecto: 12 meses

Tipo de Proyecto: Descriptiva

RESUMEN

En esta tesis se desarrolla principalmente un mecanismo formal de clasificación automática de un repositorio de objetos de aprendizaje, según el valor de un conjunto de indicadores, con el fin de mantenerlos en un ciclo permanente de mejoramiento. Esto implica construir un espacio de medida, una función de medida, una función de disimilaridad y un método de clasificación. Adicional a esto, se aplicó un proceso de medición que ayuda a dar mayor sentido a la información obtenida por cada objeto y por las posibles aglomeraciones obtenidas como producto de la clasificación.

Este mecanismo puede clasificar distintos tipos de objetos de aprendizaje, sin importar que ellos pertenezcan a categorías diferentes, esto permite que pueda ser utilizado en cualquier repositorio de objetos de aprendizaje.

CAPÍTULO 1

INTRODUCCIÓN

Es realmente importante que se logre identificar apropiadamente el problema que motiva el desarrollo de esta tesis, pues es el problema en sí el que le da sentido a ella. Como se menciona más adelante, el mecanismo que se desarrolla no garantizaría directamente la calidad de la información contenida en los objetos de aprendizaje, pero seguramente ayudaría a determinar cuáles deben pasar por mecanismos más detallados de evaluación y mejoramiento, y cuales todavía no. En todo caso, podría por lo menos mantener al educando y al educador en un "estado de alerta", que le permitiría no asumir como verdad absoluta aquello que está expuesto en los objetos de aprendizaje no certificados.

Es deseable que una vez se tenga clara la motivación de la tesis, se continúe con una mirada rápida a la literatura relacionada con el tema, para que así se pueda pasar a clarificar los conceptos teóricos en los que se basa la investigación, pues son ellos los pilares en los que se fundamenta. Entre estos conceptos teóricos se tiene: el análisis de cluster, los objetos de aprendizaje y un proceso de medición.

Una vez se tiene todo este fundamento, se motiva al lector a iniciar el recorrido por la propuesta del mecanismo formal de clasificación, el cual incluye una definición de un espacio de medida, un modelo de clasificación, un contexto abstracto, un mecanismo de valoración y un ejemplo aplicado del mecanismo de clasificación.

Finalmente esta tesis termina con unas conclusiones y un trabajo futuro que le permiten visualizar los logros alcanzados y la futura investigación a plantear.

CAPÍTULO 2

DESCRIPCIÓN DEL PROBLEMA

La manera en la que actuamos en el mundo; esto es, en nuestra vida cotidiana, en las diferentes actividades que abordamos en el día a día; depende directamente de la imagen que sobre la realidad tenemos. Aunque se habla de una distancia entre percepción y realidad, en los hechos, nuestra acción no distingue entre una y otra. Así, percepción y realidad, terminan siendo lo mismo.

Ahora bien, la manera en la que comprendemos el mundo, es decir, nuestra percepción, es determinada por las experiencias que vamos teniendo en la vida y por la manera en que nos apropiamos de ellas. Una muy importante fuente de experiencias es la escuela en todos sus niveles.

La educación, podría entenderse como el proceso mediante el cual los individuos hacen suya una cultura, esto es, se apropian de una forma característica de comprender la realidad y de actuar en ella. Obviamente que la cultura que es transmitida en los procesos educativos no es un ente fijo ni monolítico; sino que hay en ella una gran diversidad. No obstante, nadie puede trascender los límites propios que la cultura le presenta. Por tanto, cuando nos educamos recibimos una forma posible de vivir nuestra cultura.

En el presente, los procesos educativos formales de transmisión de la cultura se han visto enriquecidos por la aparición de nuevas fuentes informativas que abren, a quienes se están formando, una multiplicidad de opciones culturales. Ello ha implicado la necesidad de transformar las formas en las que se organiza la educación, para dar cabida a esa amplia variedad de formas culturales que son transmitidas, por ejemplo a través de los "objetos de aprendizaje" que circulan en la Internet. Un objeto de aprendizaje es un "recurso digital que puede ser reutilizado para ayudar en el aprendizaje"[44] para formación de sus estudiantes, profesores o administrativos.

Pero, la transformación organizativa necesaria para la inclusión de esos objetos de aprendizaje, trasciende lo meramente educativo hacia aspectos que conciernen a la administración de las instituciones de enseñanza. Las organizaciones educativas que hacen uso de estos objetos de aprendizaje, como en las universidades, deben mantener un repositorio/banco formal o informal de estos objetos para facilitar su utilización. Un ejemplo de un repositorio de objetos de aprendizaje es: “El Banco Nacional de Objetos de Aprendizaje del Ministerio de Educación Nacional de Colombia” (`http://64.76.190.172/drupalM/`).

Ahora bien, los objetos de aprendizaje como cualquier otra forma de transmisión cultural, contribuyen, como lo hemos venido afirmando, a crear en las personas una imagen consciente o inconsciente de la realidad que en ellos se trata. Esta imagen puede ser fiel o distorsionada dependiendo de la calidad del mismo; cuando por "calidad", comprendemos el grado de fidelidad con el que la información ofrecida por el objeto de aprendizaje, se apega a la realidad; aún en el entendido de que la "realidad" es algo que tiene una multiplicidad de caras.

El problema al que queremos apuntar es el siguiente: Dado que nuestra acción depende de la manera en la que comprendemos la realidad, ¿qué ocurriría si esa comprensión fuera mayoritariamente distorsionada? En otras palabras, si constantemente nos mantuviéramos recibiendo una información equivocada sobre el acontecer, nuestra acción se correspondería con el error; por tanto, tendería a ser una acción equivocada. En una sociedad en la que todo el mundo o la mayoría actúa de manera equivocada, las consecuencias serían, como lo podemos suponer, desastrosas.

Sería entonces ideal, poder disponer de un mecanismo formal de clasificación de los objetos almacenados en un repositorio, para determinar por ejemplo: ¿Cuáles de ellos deben ser mejorados? ¿Cuáles deben ser descartados? ¿Cuáles son los mejores objetos que se disponen sobre cierto tema? ¿Cuáles no lo son?

Sin embargo, una mirada rápida al “Banco Nacional de Objetos de Aprendizaje del Ministerio de Educación de Colombia” no evidencia la existencia de un mecanismo de clasificación. Por supuesto, un mecanismo así no garantizaría directamente la calidad de la información contenida en los objetos, pero seguramente ayudaría a determinar cuáles deben pasar por mecanismos más detallados de evaluación y mejoramiento, y cuales todavía no. En todo caso, podría por lo menos mantener al educando y al educador en un "estado de alerta", que le permitiría no asumir como verdad absoluta aquello que está expuesto en los objetos de aprendizaje no certificados.

Por otra parte, mantener en un ciclo de mejoramiento continuo de los objetos que salieron clasificados en una categoría no muy buena, permitirá seguramente que las personas o usuarios de estos repositorios cuenten con objetos cada vez mejores, disponiendo de muy buenos objetos de aprendizaje lo que ayudaría por ejemplo a:

- Que los usuarios de estos objetos se formen una visión menos distorsionada de la realidad, disminuyendo así el impacto que estos objetos de baja calidad ejercerían sobre el comportamiento individual o colectivo.

- Crear un ambiente de mayor confianza en el uso de los objetos de aprendizaje y en las habilidades o conceptos generados a partir de su utilización.
- Focalizar recursos tanto humanos como económicos en mejorar los actuales "peores objetos" del repositorio.

Por lo tanto la formulación del problema de investigación es:

¿Es posible definir un mecanismo formal de clasificación automática de un repositorio de objetos de aprendizaje, a partir de una valoración individual de los objetos, con el fin de mantenerlos en un ciclo permanente de mejoramiento?

Por supuesto, lo ideal de este mecanismo formal de clasificación es que no dependa del significado de lo que se evalué en los objetos, sino del valor asignado a ellos. Este mecanismo debe operar indistintamente a los indicadores evaluados; así que es posible que dos instituciones valoren el mismo objeto a través de distintos grupos de indicadores. Al mismo tiempo que no es necesario que objetos de diferente tipo deban ser evaluados usando los mismo indicadores.

Si todos los objetos del repositorio se evaluaran con los mismo indicadores, entonces, una forma de solucionar el problema se encuentra en la literatura [1] y [2] y es haciendo aglomeración (Clustering), el cual es el proceso de agrupar datos en clases (cluster) o aglomeraciones, en donde un objeto dentro de la aglomeración tiene alta similaridad con los otros objetos de la misma aglomeración; pero a la vez tiene alta disimilaridad con objetos de otras aglomeraciones. Sin embargo, suponer que todos los objetos de aprendizaje almacenados en un repositorio son del mismo tipo, o que sólo se puede evaluar aquellas cosas comunes a todos, no es algo muy realista. Por tanto, esta estrategia de solución es valida sólo para casos especiales o muy particulares. Por otro lado, no se ha encontrado hasta ahora en la literatura analizada, ni en conversaciones con expertos, un mecanismo de Clustering ya definido que permita clasificar elementos evaluados con diferentes indicadores.

Por supuesto que existen otras formas para clasificar los objetos de aprendizaje como la comprensión semántica de textos [42], la clasificación documentos Web [16], la similaridad de imágenes [35], o por conjuntos frecuentes (frequent itemsets) [4] y [11]. Se podría continuar así por mucho tiempo, citando uno y otro artículo que mencione algoritmos o criterios diversos de clasificación, para clasificar diversos tipos de objetos, ya sean documentos, imágenes, incluso archivos de sonido. De nuevo entre toda la bibliografía revisada, no se encontró ninguna que resolviera el problema aquí planteado.

Finalmente, se espera con esta investigación, dar respuesta al problema de investigación, contribuyendo así a disminuir el impacto que objetos de baja calidad almacenados en los repositorios ejercen sobre el comportamiento individual o colectivo en todo el mundo.

CAPÍTULO 3

OBJETIVOS

Como ya se ha indicado, la manera en que actuamos depende de la imagen de la realidad que tenemos. Esta imagen, puede verse afectada por el uso de los objetos de aprendizaje que cada vez son más populares, estos objetos, contribuyen a crear una imagen que puede ser exacta o distorsionada dependiendo de su calidad, con respecto a la veracidad de la información contenida en ellos. Por tanto, si se quiere ayudar a disminuir el impacto que objetos de baja calidad ejercen sobre el comportamiento individual o colectivo, sería muy útil poder clasificarlos. Por lo anterior, los objetivos de esta tesis son:

3.1. Objetivo general

Desarrollar un mecanismo formal de clasificación automática de un repositorio de objetos de aprendizaje, a partir de una valoración individual de los objetos, con el fin de mantenerlos en un ciclo permanente de mejoramiento.

Ahora, para poder desarrollar un mecanismo formal de clasificación, es indispensable construir un espacio de medida que sirva para clasificar los objetos, al igual que un método de clasificación que utilice dicho espacio, por lo anterior, los objetivos específicos son:

3.2. Objetivos específicos

1. Construir un espacio de medida (la terna (Conjunto de objetos X, la sigma álgebra Σ, una función de medida μ)) que sirva para clasificar automáticamente un conjunto de objetos de aprendizaje según la valoración individual de los objetos.

2. Definir el método de clasificación que utilice la medida definida.

3. Determinar los indicadores para clasificar los objetos de aprendizaje en un contexto particular.

4. Establecer un mecanismo de valoración de los indicadores definidos.

5. Desarrollar un ejemplo del método de clasificación para el contexto particular definido.

CAPÍTULO 4

ESTADO DEL ARTE

Muy pocas personas en el mundo podrían pensar que la calidad, en general, no es importante, más aún, muy pocas personas pensarían que la calidad es algo que se consigue de forma instantánea. Casi de manera generalizada se acepta que la calidad es un proceso continuo, lento y permanente. También se acepta que uno de los principales enemigos de la calidad es la complacencia y es ahí en donde tiene lugar este trabajo de investigación.

En el mundo existen muchas colecciones de objetos de aprendizaje, es más, existen repositorios o bancos de objetos de aprendizaje como el "Banco Nacional de Objetos de Aprendizaje del Ministerio de Educación de Colombia", en ellos, es común observar mecanismos de búsqueda y seguramente mecanismos de clasificación sobre ciertas temáticas particulares. Sin embargo, una vez un objeto es ingresado al repositorio, éste permanece constante y a la espera de ser consultado, tal y como permanece un libro en una biblioteca.

En el documento "*Quantitative Analysis of Learning Object Repositories*"[32] se presenta un análisis cuantitativo sobre algunos de los repositorios de objetos de aprendizaje, analizando por ejemplo el tamaño, las contribuciones y la popularidad. Entre los repositorios estudiados en este documento se encuentran: *Ariadne*, *Merlot*, *BioDiTRL*.

Algunos repositorios pueden contar con algún tipo de herramienta de evaluación, tal y como se evidencia en "*Web-Based Tools for Learning Object Evaluation*"[25], el cual presenta una revisión a los mecanismos de evaluación de los objetos de aprendizaje realizada en: *CLOE*, *MERLOT* y *DLNET*. Adicionalmente el documento introduce a *eLera*, el cual es un conjunto de herramientas basadas en Web, especialmente diseñadas para comunicar a profesores, estudiantes, diseñadores instruccionales y desarrolladores. También esta diseñada para ayudar a investigadores a capturar datos sobre procesos de evaluación y percepción de la calidad de objetos de aprendizaje.

Este documento se complementa con "*Web-Based Tools for Collaborative Evaluation of Learning Resources*"[26], debido a que este último detalla el elemento evaluación colaborativa ilustrado en [25].

Un punto a resaltar es que los resultados de los mecanismos de búsqueda pueden ser manipulados, reordenados o priorizados según criterios ajenos a un sistema formal, ya sea por cuestiones políticas, económicas o sociales. Estas posibles manipulaciones han logrado despertar preocupaciones en algunas personas como Micah White (activista independiente que escribe en Adbusters Magazine[1].) quien piensa que "quien controla los resultados de la búsqueda controla también al autor de la búsqueda", pues visualiza un futuro en donde la diversidad de las ideas desaparezca y "todos" pensemos y en últimas actuemos tal y como quieren los que controlan los motores de búsqueda como Google.

Sin lugar a dudas, existe innumerable bibliografía sobre calidad, calidad total, procedimientos de clasificación, procedimientos de aglomeración, objetos de aprendizaje, repositorios de objetos de aprendizaje, por tanto y después de realizar una búsqueda sistemática de artículos actualizados y documentos en general sobre las temáticas relacionadas, se han podido extraer las siguientes referencias:

En el artículo "*Survey of clustering data mining techniques*" [2] se puede ver un estudio de las técnicas de agrupación de *Data Mining*, al profundizar en una clasificación de los algoritmos de aglomeramiento (*Cluster*) más conocidos, sin intentar ser exhaustiva. Este artículo toma un interés significativo para esta investigación, pues los algoritmos citados, pueden ser utilizados de forma general en distintos ámbitos, como en este caso en la clasificación de objetos de aprendizaje.

Existen sistemas más especializados como por ejemplo sistemas para clasificar documentos o páginas Web, como lo muestra [16] en su artículo "*Efficient phrase-based document indexing for Web document clustering*". En él, los autores presentan un sistema compuesto por cuatro componentes: identificación de la estructura del documento, representación gráfica del índice del documento, cálculo de similaridad entre documentos y cluster incremental, en un intento de mejorar la solución al problema de la agrupación de documentos en un dominio web. Ellos indican que las aplicaciones potenciales incluyen: la agrupación automática de resultados de un motor de búsqueda, la recuperación de la información, la medición de la similitud de documentos, entre muchos otros. Como se puede apreciar entre las aplicaciones se encuentran la similaridad de documentos, estos sistemas y similares, analizan el contenido del documento y usando algún criterio miden su similaridad con otros documentos, y así logran aglomerarlos según el grado de similaridad encontrados.

En el documento *"Learning to Rank Relational Objects and Its Application to Web Search*"[36] se propone un nuevo método para el desempeño de la tarea de optimización particularmente en una implementación basada sobre SVM. Este método puede ser aplicado a una gran variedad de problemas en la búsqueda Web.

[1] `www.micahmwhite.com`

En "*Learning to Rank: From Pairwise Approach to Listwise Approach*"[5] se propone un nuevo método probabilístico para clasificar (*ranking*) objetos en general (*ListNet*), el cual, según sus creadores, es mejor que la aproximación tradicional, éste hace uso de un modelo de red neuronal y utiliza para optimizar el algoritmo del gradiente descendente. También muestra algunos de los resultados obtenidos en distintos experimentos. Este tipo de métodos son importantes para el desarrollo de aplicaciones como la recuperación de documentos.

En el documento *"Finding The Most Probable Ranking of Objects with Probabilistic Pairwise Preferences"*[34] se estudia el problema de encontrar la clasificación más probable de un conjunto de objetos cuando son conocidas preferencias probabilísticas para cada par de objetos. Utilizan la teoría de grafos para resolver el problema y proponen tres algoritmos para encontrar la clasificación más probable.

La presentación "*Learning to Rank:Problem, Challenge, and Opportunity*"[21], hace un recorrido rápido por el problema de clasificación (*ranking*), específicamente en la clasificación de documentos. En el documento se presentan someramente los diferentes modelos desde los tradicionales, hasta los probabilísticos. A pesar de su aparente sencillez, brinda un panorama interesante sobre la problemática y sus enfoques.

El documento *"Learning to Rank for Information Retrieval Using Genetic Programming"*[46] presenta un método llamado "*RankGP*", el cual se propone como una efectiva función de clasificación (*Ranking*) que hace uso de los algoritmos genéticos. Esta función puede combinar diferentes tipos de características, por ejemplo: del contenido, de la estructura y características de consultas independientes.

En el artículo "*Learning Objects Clustering based on Semantic Understanding of Text*" [42], se puede observar otro sistema que mide la similaridad entre elementos, pero esta vez sobre objetos de aprendizaje basados sobre la comprensión de los textos internos al objeto de aprendizaje. Ellos indican que la calidad de la agrupación lograda con este modelo supera el modelo tradicional de espacios vectoriales.

No siempre se necesita medir la similaridad entre documentos, sino también sobre imágenes y este es el caso que se observa en el artículo "*Learning object representations by clustering banana wavelet responses*" [35], en el cual se introduce un sistema que aprende como representar un objeto a partir de imágenes de entrenamiento. Cada imagen es pre-procesada usando *Banana Wavelets*, el resultado es una representación del objeto por medio de líneas curvas. Estas líneas se insertan en un algoritmo de clasificación que aprende a separar las características específicas del objeto.

En el documento "*Large Scale Online Learning of Image Similarity Through Ranking*"[7] se presenta *OASIS* (*Online Algorithm for Scalable Image Similarity*) un algoritmo eficiente que permite crear aplicaciones importantes como la búsqueda de imágenes que son similares a una imagen dada o aplicaciones que buscan videos que son relevantes a un video dado. En el documento se detalla un poco sobre la función de similaridad relativa a un par de imágenes y se mencionan algunos de los experimentos realizados, al igual que los resultados obtenidos.

En el documento *"Towards to an automatic semantic annotation for multimedia learning objects*"[37], los autores describen y evalúan una nueva aproximación (algoritmo) para generar una anotación semántica para recursos multimedia, con el fin de poder realizar búsquedas.

En "*TVGraz: Multi-Modal Learning of Object Categories by Combining Textual and Visual Features*"[20] presentan un conjunto de datos público, llamado *TVGraz* con más de 4030 imágenes asociadas a páginas web y agrupadas en 10 categorías, para probar el desempeño de algoritmos de categorización de objetos visuales basados en Web que utilice tanto información textual como visual. Además presentan un kernel que provee solución natural a el problema.

Como se ha podido observar hasta ahora, la clasificación se ha realizado usando como insumo el contenido de los documentos o de las imágenes, sin embargo en muchas ocasiones, se necesitan clasificar elementos por el valor de un conjunto de elementos (*frequent itemsets*), como por ejemplo el valor obtenido por un objeto de aprendizaje en un conjunto de indicadores.

El artículo "*Using Conjunction of Attribute Values for Classification*"[11] presenta una serie de algoritmos de clasificación que usan "*frequent itemsets*". Otro artículo es "*Large hierarchical document clustering using frequent itemsets*" [4] el cual presenta un nuevo enfoque para abordar la clasificación de documentos con requisitos especiales, tales como: la dimensionalidad, el alto volumen, y la facilidad de la navegación significativa por los clusters. Ellos definen un criterio de agrupación usando algunas palabras comunes.

Adicional a todo esto existen documentos como: "*Proceedings of the NIPS 2005 Workshop on Learning to Rank*"[28], que agrupa información relacionada, por ejemplo en este *Workshop* se presentan siete artículos relacionados con la clasificación (*ranking*).

En el documento "*Outlier Detection and Ranking Based on Subspace Clustering*"[39], se describe brevemente un modelo consistente para clasificar *outliers* en la presencia de diferentes tipos de atributos para datos con muchas dimensiones, llamado "*OutRank*".

Por otro lado, el documento "Calidad en objetos de aprendizaje"[43] propone un sistema para medir la calidad de un objeto de aprendizaje OA [2]basados en el estándar ISO 9126 (*Information technology- Software Product Evaluation-Quality characteristics and guidelines for their use*). El documento es importante para esta tesis, pues propone un conjunto de características y de rasgos a medir en un objeto de aprendizaje, que cubre características de: Funcionalidad, Usabilidad, Eficiencia y de Reutilización.

En el documento "*Learning User Preferences for Sets of Objects*"[12] se presenta un método para aprender las preferencias de los usuarios sobre un conjunto de objetos y muestran que es efectivo tanto para problemas teóricos como del mundo real.

[2]entendido para ellos como todo recurso digital, independiente y reutilizable, deseablemente interactivo y multimedia, creado con la expresa intención de apoyar a la instrucción y al aprendizaje.

La presentación "*Learning to Rank: from Pairwise Approach to Listwise Approach*"[22] brinda una excelente panorámica de los distintos métodos y modelos relacionados con la clasificación (*ranking*) en aplicaciones de recuperación de información. Este documento es clave para vislumbrar la magnitud y complejidad que se encuentra involucrada en la clasificación (*ranking*) de elementos, para el caso de documentos. Además permite comprender muchos de los otros documentos citados en esta tesis.

En "*Metrics for Learning Object Metadata*"[30] se presenta una investigación que se enfoca en la búsqueda de métricas en tres principales campos de los objetos de aprendizaje (calidad, la similaridad y la relevancia). También expone una metodología para explorar nuevas áreas donde las métricas pueden ser usadas.

En el documento "*Learning to combine distances for complex representations*"[45] se presenta un marco de trabajo (*framework*) que permite combinar diferentes representaciones complejas de un problema de aprendizaje dado y/o definir disimilaridades diferentes sobre estas representaciones, todo con el fín de clasificar objetos complejos.

En *"Object Level Ranking: Bringing Order to Web Objects*"[27] se introduce *PopRank*, un modelo de análisis de un rango de objetos sin un dominio específico. Estudia como calcular la popularidad de un objeto *Web* almacenado basado en la relación con los otros objetos.

En "*Learning sparse metrics via linear programming*"[38] se propone un método para construir distancias relativas preservando el mapeo en bajas dimensiones (espacio de mapeo). Este método permite conocer una función de distancia desconocida, o aproximarla a una función conocida, con la adición de propiedades de reducción de la distancia en tiempo computacional. La formulación está basada en la solución de un problema de optimización de programación lineal que encuentra un mapeo óptimo para el conjunto de datos dados y la relación de distancia. Desarrolla una nueva aproximación para las funciones de distancia desde un conjunto de relaciones de distancia relativa.

En*"Use of contextualized attention metadata for ranking and recommending learning objects*"[31] se propone y detalla el uso de metadatos de atención contextual para crear clasificación (*ranking*) y recomendar métricas que involucren la experiencia del usuario.

Finalmente, si lo que se desea es tener un repositorio de objetos de aprendizaje de calidad, es necesario entrar en el ciclo continuo de mejoramiento, es decir, buscar los objetos del repositorio que merecen ser mejorados, extraerlos, mejorarlos e ingresarlos nuevamente al repositorio, reclasificándolos y encontrando los nuevos objetos a mejorar, y así sucesivamente... Por otro lado, uno de los grandes enemigos de la calidad es la falta de constancia, pues los movimientos de mejoramiento de calidad son procesos de largo plazo.

Este movimiento hacia la calidad total, no se trata únicamente de mejorar lo que está mal, se trata también de encontrar sus causas para así poder eliminar el problema de raíz. Y es en este punto en donde los procesos de clasificación y aglomeración (*cluster*) tienen verdadera importancia, pues una vez clasificados los objetos de aprendizaje, se pueden identificar no

sólo los objetos a mejorar, sino grupos de objetos y determinar las características que son comunes en ellos y así poder tratar de encontrar las posibles causas que provocan esa falta de calidad en el repositorio.

En toda la consulta bibliográfica realizada no se ha encontrado bibliografía que combine estas temáticas para resolver el problema ya mencionado. Sí se ha encontrado bibliografía que hable sobre clasificación de documentos, que en últimas tiene algo de similar al contexto general de lo ya expuesto.

CAPÍTULO 5

MARCO TEÓRICO

En este marco teórico se desarrolla una visión panorámica de tres temas fundamentales para esta tesis como lo son: El análisis de cluster (aglomeración), los objetos de aprendizaje y el proceso de medición.

En el primer tema, el análisis de cluster, se habla de los requerimientos de un algoritmo de aglomeración, de los tipos de datos utilizados para calcular las matrices de similaridad o disimilaridad y de los métodos de clasificación de dichos algoritmos. Este tema es de vital importancia para desarrollar la Sección 6.1 en la página 65 y la Sección 6.2 en la página 68.

En el segundo tema, los objetos de aprendizaje, se habla de una taxonomía de los tipos de objetos de aprendizaje, de sus características y de los metadatos utilizados para describirlos. Este tema se necesita para desarrollar parte de la Sección 6.1 en la página 65 y de la Sección 6.3 en la página 71.

El tercer y último tema: el proceso de medición, se habla sobre una ontología sobre la medición basada en diferentes estándares, en la cual se definen sub-ontologías como: Modelo de calidad, Medida, Formas de medir y Acción de medir. Esta ontología es fundamental para desarrollar la Sección 6.3 en la página 71 y Sección 6.4 en la página 76.

5.1. Análisis de Cluster

Basados en lo que menciona Jiawei Han [15], se puede decir en esta sección que:

Aglomerar (Clustering) es el proceso de agrupar datos en clases (cluster) o aglomeraciones, en donde un objeto dentro de la aglomeración tiene alta similaridad con los otros objetos de

la misma aglomeración pero a la vez tiene alta disimilaridad con objetos de otras aglomeraciones.

Existen muchas técnicas de aglomeración las cuales se pueden organizar en las siguientes categorías: métodos de partición, métodos jerárquicos, métodos basados en densidad, métodos basados en grillas, métodos basados en modelos, métodos basados en patrones frecuentes y aglomeración.

No hay duda que el análisis de clases (cluster) es importante en la actividad humana, ya que nosotros aprendemos desde muy jóvenes cómo distinguir entre los diferentes objetos entre sí, ya sean físicos o abstractos y es esa capacidad la que nos permite interactuar con el mundo que nos rodea, permitiéndonos hacer por ejemplo: investigación de mercadeo, reconocimiento de patrones, análisis de datos, procesamiento de imágenes, realizar taxonomía de plantas y animales, entre muchas otras actividades.

La investigación en el análisis de clases tiene un gran desarrollo gracias a las contribuciones de diferentes áreas como: la biología, las bases de datos (la minería de datos), la inteligencia artificial (el aprendizaje de máquinas), la estadística, el mercadeo.

Como una rama de la estadística, el análisis de clases ha sido ampliamente estudiado por muchos años, enfocándose principalmente sobre análisis de clases basados en distancia, usando por ejemplo el método de k-medias (k-means).

Existe una serie de requerimientos sobre los algoritmos de aglomeración:

- **Escalabilidad:** Trabajar con un número alto de objetos.
- **Trabajar con diferentes tipos de datos:** Operar con diferentes tipos, como por ejemplo datos: basados en intervalos, binarios, nominales, ordinales y una combinación de ellos.
- **Descubrir agrupaciones en formas arbitrarias:** Ya que la mayoría de los algoritmos de clasificación se basan en medidas de distancia Euclidiana o Manhanttan (la del taxista), las cuales producen sólo formas esféricas.
- **Mínima cantidad de parámetros de entrada:** Lo ideal de un algoritmo sería no tener o tener la mínima cantidad de parámetros que puedan afectar la aglomeración.
- **Habilidad para trabajar datos con ruido:** Se pueda realizar la aglomeración aún si en el conjunto de datos a clasificar hay por ejemplo: datos erróneos o incompletos.
- **Aglomeración incremental e insensibilidad a la cantidad de objetos de entrada:** Disponer de un algoritmo que no dependa de la cantidad de objetos iníciales o que se puedan ir agregando o modificando los mismo sin mayores dificultades.

Cuadro 5.1: Matriz de datos

$Objeto/Atributo$	$Atributo_1$	$Atributo_2$	$\cdots$	$Atributo_p$
$Objeto_1$	a_{11}	a_{12}	$\cdots$	a_{1p}
$Objeto_2$	a_{21}	a_{22}	$\cdots$	a_{2p}
$Objeto_3$	a_{31}	a_{32}	$\cdots$	a_{3p}
$\cdots$	$\cdots$	$\cdots$	$\ddots$	$\cdots$
$Objeto_n$	a_{n1}	a_{n2}	$\cdots$	a_{np}

- **Alta dimensionalidad:** Los objetos como ya se ha dicho, pueden estar representados o definidos por una cantidad de atributos o dimensiones, por tanto sería ideal poder trabajar con un número elevado de dimensiones por objeto.
- **Agrupación basada en restricciones:** Disponer de algoritmos que creen aglomeraciones que satisfagan ciertas restricciones o condiciones especificadas.
- **Interpretabilidad y usabilidad:** Que los resultados de las aglomeraciones puedan ser fácilmente interpretados, comprendidos y que sean sobre todo usables.

Ahora se va a proceder a estudiar los tipos de datos empleados en el análisis de clases.

5.1.1. Tipos de datos en el análisis de clases

Supongamos que se tienen n objetos que representan por ejemplo: páginas web, aplicaciones multimedia, o bien, sistemas de aprendizaje, o tal vez quizás, representan casas, personas, vehículos, entre muchos otros tipos de objetos. Y cada uno de ellos se encuentra descrito a través de p atributos como número de enlaces, cantidad de imágenes, tipos de idiomas, peso, edad y así sucesivamente. Estos objetos se pueden representar por medio de una matriz, de dimensión $n \dot{\times} p$, conocida como matriz de datos o estructura objeto-por-variable (ver Tabla 5.1).

Un procedimiento habitual antes de aplicar un algoritmo de aglomeración es utilizar esta matriz para determinar la distancia entre cada par de objetos, creando así otra matriz, de dimensión n x n, llamada matriz de disimilaridad o estructura objeto-por-objeto (ver Tabla 5.2).

En donde d(i,j) mide la diferencia o disimilaridad entre el objeto i y el objeto j. Esta función que mide la disimilaridad debe garantizar que [1]:

- $d(i, j)$, es un número positivo o cero. Este valor estará muy cercano a cero si el objeto i y el j son muy similares o cercanos uno del otro, o será un valor mayor en caso contrario.

Cuadro 5.2: Matriz de disimilaridad

$Objeto/Objeto$	$Objeto_1$	$Objeto_2$	$Objeto_2$	$\cdots$	$Objeto_p$
$Objeto_1$	0				
$Objeto_2$	$d(2,1)$	0			
$Objeto_3$	$d(3,1)$	$d(3,2)$	0		
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\ddots$	
$Objeto_p$	$d(p,1)$	$d(p,2)$	$d(p,3)$	$\cdots$	0

- $d(i,j) = d(j,i)$, la función de disimilaridad debe ser simétrica, es decir la disimilaridad entre el objeto i y el objeto j debe ser igual que entre el objeto$j$ y el objeto i.
- $d(i,i) = 0$, la disimilaridad entre un objeto y él mismo siempre debe ser cero.
- $d(i,j) \leq d(i,k) + d(k,j)$, la disimilaridad entre el objeto i y el objeto j debe ser menor o igual a la suma de la disimilaridad entre el objeto i y k con la disimilaridad entre el objeto k y j, esta propiedad se conoce como desigualdad triangular.

La matriz de disimilaridad que satisface estas restricciones es una matriz simétrica, cuya diagonal principal es cero. Por ello, podemos escribir esta matriz como una matriz triangular, para el ejemplo una matriz triangular inferior.

Dependiendo de la naturaleza de los atributos (por ejemplo: basados en intervalos, binarios, nominales, ordinales y una combinación de ellos) la función de disimilaridad se calcula de forma diferente, tal y como se estudia a continuación.

5.1.1.1. Variables basadas en intervalos

Entre las medidas utilizadas con estas variables se encuentran [1] la métrica ***Minkowski***, ver la ecuación 5.5:

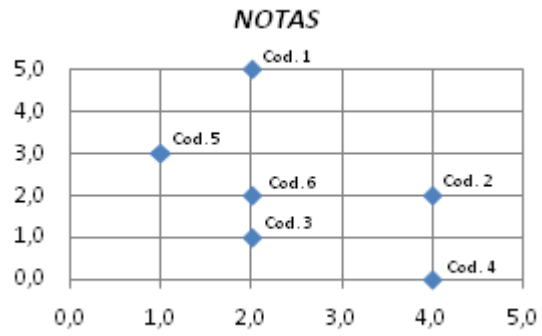
$$d(i,j) = \left[\sum_{i=1}^{n} |x_{ij} - x_{ik}|^p \right]^{1/p} ; p \geq 1 \tag{5.1}$$

y sus casos especiales:

- $p = 1$, llamada ***Métrica Manhanttan*** (o del taxista), ver la ecuación 5.2:

$$d(i,j) = \left[\sum_{i=1}^{n} |x_{ij} - x_{ik}| \right] \tag{5.2}$$

Figura 5.1: Ejemplo - Variables basadas en intervalos



- $p = 2$, llamada ***Métrica Euclidiana,*** ver la ecuación ver la ecuación 5.3:

$$d(i,j) = \left[\sum_{i=1}^{n} |x_{ij} - x_{ik}|^2\right]^{1/2} \equiv \sqrt{\sum_{i=1}^{n} |x_{ij} - x_{ik}|^2} \tag{5.3}$$

- $p = \infty$, llamada ***Métrica Chebychev***, ver la ecuación 5.4:

$$d(i,j) = \lim_{p\to\infty} \left[\sum_{i=1}^{n} |x_{ij} - x_{ik}|^p\right]^{1/p} \equiv \max_{i=1,\ldots,n} |x_{ij} - x_{ik}| \tag{5.4}$$

Un ejemplo hipotético con las notas de seis estudiantes, pueden servir para ilustrar el manejo de este tipo de variables, ver Tabla 5.3 y Figura 5.1. Al utilizar estos datos para calcular la disimilaridad entre los estudiantes, se obtiene por cada tipo de medida una matriz (ver Tabla 5.4).

Cuadro 5.3: Ejemplo - Variables basados en intervalos

$CódigoEstudiante$	$Nota_1$	$Nota_2$
1	2,0	5,0
2	4,0	2,0
3	2,0	1,0
4	4,0	0,0
5	1,0	3,0
6	2,0	2,0

Cuadro 5.4: Distacias sobre el ejemplo de variables basadas en intervalos

Manhanttan	1	2	3	4	5	6
1	0,0	–	–	–	–	–
2	5,0	0,0	–	–	–	–
3	4,0	3,0	0,0	–	–	–
4	7,0	2,0	3,0	0,0	–	–
5	3,0	4,0	3,0	6,0	0,0	–
6	3,0	2,0	1,0	4,0	2,0	0,0

Euclidiana	1	2	3	4	5	6
1	0,0	–	–	–	–	–
2	3,6	0,0	–	–	–	–
3	4,0	2,2	0,0	–	–	–
4	5,4	2,0	2,2	0,0	–	–
5	2,2	3,2	2,2	4,2	0,0	–
6	3,0	2,0	1,0	2,8	1,4	0,0

Chebychev	1	2	3	4	5	6
1	0,0	–	–	–	–	–
2	3,0	0,0	–	–	–	–
3	4,0	2,0	0,0	–	–	–
4	5,0	2,0	2,0	0,0	–	–
5	2,0	3,0	2,0	3,0	0,0	–
6	3,0	2,0	1,0	2,0	1,0	0,0

Existen variantes asociadas ***Métrica Minkowski***, ver ecuación 5.5, en las que a cada variable se le asigna un peso o porcentaje de importancia (w_i).

$$d(i,j) = \left[\sum_{i=1}^{n} w_i \left| x_{ij} - x_{ik} \right|^p \right]^{1/p} ; p \geq 1 \tag{5.5}$$

Se puede asumir que en la métrica original de ***Minkowski***, ver ecuación 5.1, cada w_i equivale a uno (1).

5.1.1.2. Variables binarias

En algunas ocasiones es necesario trabajar con variables binarias (Sí o No, Verdadero o Falso, 1 ó 0, etc), pero a pesar de que sólo pueden tomar dos valores, es posible que no sea de igual importancia que tome uno u otro valor. Es decir, no da igual por ejemplo que una persona esté o no este infectada por algún tipo de virus biológico. Por tanto tenemos, Variables binarias simétricas y variables binarias asimétricas.

Para medir la similaridad o la disimilaridad entre variables binarias es necesario comprender la tabla 5.5.

Cuadro 5.5: Variables binarias

		$objeto_j$		
		1	0	$suma$
$objeto_i$	1	q	r	$q+r$
	0	s	t	$s+t$
	$suma$	$q+s$	$r+t$	p

En donde:

- q: Número de atributos cuyo valor corresponde a 1 en ambos objetos.
- r: Número de atributos cuyo valor es 1 en el $objeto_i$. y 0 en el $objeto_j$.
- s: Número de atributos cuyo valor es 0 en el $objeto_i$. y 1 en el $objeto_j$.
- t: Número de atributos cuyo valor corresponde a 0 en ambos objetos.
- p: Número total de atributos por cada objeto.

Para calcular la disimilaridad entre variables binarias que son simétricas se utiliza la ecuación 5.6.

$$d(i,j) = \frac{r+s}{q+r+s+t} = \frac{r+s}{p}, \tag{5.6}$$

pero si lo que se desea es calcular es la similaridad, basta con aplicar la ecuación 5.7.

$$sim(i,j) = 1 - d(i,j). \tag{5.7}$$

Por otro lado, si las variables binarias son asimétricas, se debe utilizar la ecuación 5.8.

$$d(i,j) = \frac{r+s}{q+r+s}, \tag{5.8}$$

pero si lo que se desea es calcular es la similaridad, basta con aplicar la misma ecuación 5.7, salvo que el valor de la disimilaridad debe ser el calculado con la ecuación 5.8.

Un nuevo ejemplo hipotético con cuatro atributos binarios de seis estudiantes, pueden servir para ilustrar el manejo de este tipo de variables, ver Tabla 5.6. Al utilizar estos datos para calcular la disimilaridad y la similaridad, entre los estudiantes, se obtienen una matriz por cada combinación (ver Tabla 5.7).

Cuadro 5.6: Ejemplo de variables binarias

$CódigoEstudiante$	$Atributo_1$	$Atributo_2$	$Atributo_3$	$Atributo_4$
1	0	0	0	1
2	1	0	0	0
3	1	1	1	0
4	1	0	1	0
5	1	0	1	1
6	0	1	1	0

Cuadro 5.7: Distacias sobre el ejemplo de variables binarias

Disimilaridad - Variables binarias simétricas

	1	2	3	4	5	6
1	0,00	–	–	–	–	–
2	0,50	0,00	–	–	–	–
3	1,00	0,50	0,00	–	–	–
4	0,75	0,25	0,25	0,00	–	–
5	0,50	0,50	0,50	0,25	0,00	–
6	0,75	0,75	0,25	0,50	0,75	0,00

Disimilaridad - Variables binarias asimétricas

	1	2	3	4	5	6
1	0,00	–	–	–	–	–
2	1,00	0,00	–	–	–	–
3	1,00	0,67	0,00	–	–	–
4	1,00	0,50	0,33	0,00	–	–
5	0,67	0,67	0,50	0,33	0,00	–
6	1,00	1,0	0,33	0,67	0,75	0,00

Similaridad - Variables binarias simétricas

	1	2	3	4	5	6
1	1,00	–	–	–	–	–
2	0,50	1,00	–	–	–	–
3	0,00	0,50	1,00	–	–	–
4	0,25	0,75	0,75	1,00	–	–
5	0,50	0,50	0,50	0,75	1,00	–
6	0,25	0,25	0,75	0,50	0,25	1,00

Similaridad - Variables binarias asimétricas

	1	2	3	4	5	6
1	1,00	–	–	–	–	–
2	0,00	1,00	–	–	–	–
3	0,00	0,33	1,00	–	–	–
4	0,00	0,50	0,67	1,00	–	–
5	0,33	0,33	0,50	0,67	1,00	–
6	0,00	0,00	0,67	0,33	0,25	1,00

Sólo por citar algunas interpretaciones:

- El objeto de código 1 y el de código 3 son totalmente diferentes.
- El objeto de código 1 y el de código 4 son totalmente diferentes, si se están utilizando valores asimétricos, pues en otro caso tienen un grado de similaridad de un 0.25 (25 %).

5.1.1.3. Variables nominales

Las variables nominales, son variables cualitativas. Estas pueden ser consideradas una generalización de las variables binarias simétricas, esto es, las variables nominales pueden tener dos o más categorías. Por ejemplo los deportes favoritos, el sexo, entre otras.

Los valores de las categorías pueden ser: letras, símbolos o un conjunto de enteros. Sin embargo estos enteros no representan un orden.

La disimilaridad entre dos objetos i y j puede ser calculada con base a la ecuación 5.11.

$$d(i,j) = \frac{p-m}{p} \tag{5.9}$$

En donde:

- m: representa el número de coincidencias.
- p: representa el número total de variables.

Es posible asignar un peso para incrementar el efecto de m.

El ejemplo que aparece en la Tabla 5.8, permite ilustrar la forma de calcular la disimilaridad (ver Tabla 5.9) . De nuevo si se desea calcular la similaridad, basta con restarle a uno el valor de disimilaridad.

Cuadro 5.8: Ejemplo de variables nominales

$CódigoEstudiante$	$Atributo_1$	$Atributo_2$	$Atributo_3$	$Atributo_4$
1	a	u	o	e
2	u	u	o	u
3	e	i	u	e
4	i	u	o	o
5	i	e	a	a
6	i	o	u	u

Cuadro 5.9: Distancias sobre el ejemplo de variables nominales

	1	2	3	4	5	6
1	0,00	–	–	–	–	–
2	0,75	0,00	–	–	–	–
3	0,75	1,00	0,00	–	–	–
4	0,50	0,50	1,00	0,00	–	–
5	1,00	1,00	1,00	0,75	0,00	–
6	1,00	0,75	0,75	0,75	0,75	0,00

5.1.1.4. Variables ordinales

Es una variable cualitativa y a diferencia de las nominales, sus categorías tienen un orden, por ejemplo: Malo, Regular, Aceptable, Bueno y Excelente; o por ejemplo: Bronce, Plata y Oro.

La disimilaridad entre el $objeto_i$ y el $objeto_j$ puede ser calculada con base al siguiente procedimiento:

1. Enumerar los posibles valores de la variable f de 1 a un valor M_f, en donde M_f representa la posición del último valor de la variable f en la respectiva escala. Por ejemplo, si los posibles valores son Bronce, Plata y Oro, enumérelos como 1, 2 y 3 en donde 1 representa Bronce, 2 Plata y 3 Oro. M_f es entonces igual a 3.

2. Convertir estos valores a un valor entre cero y uno, en donde 0.0 representa, para el ejemplo Bronce, 0.5 representa Plata y 1.0 representa Oro. Para lograrlo, es necesario realizar dos actividades:

   a) Reemplazar el valor de cada variable ordinal (r_{if}) en la matriz de datos por su respectivo valor en la escala.

   b) Aplicarle la formula 5.10.

$$v_{if} = \frac{r_{if} - 1}{M_f - 1} \tag{5.10}$$

3. Trabajarla como una variable definida en un intervalo, usando por ejemplo uno de los casos particulares de ***Minkowski*** 5.1 en la página 40.

Para comprender mejor este caso, supongamos el siguiente ejemplo hipotético, ver Tabla 5.10.

Cuadro 5.10: Ejemplo de variables ordinales

$CódigoEstudiante$	$Atributo_1$	$Atributo_2$	$Atributo_3$	$Atributo_4$
1	Excelente	Aceptable	Aceptable	Oro
2	Regular	Bueno	Regular	Oro
3	Malo	Malo	Malo	Bronce
4	Malo	Bueno	Malo	Plata
5	Bueno	Regular	Bueno	Oro
6	Malo	Regular	Regular	Plata

Se aplica el paso 1 del procedimiento, enumerar los posibles valores de la variable f de 1 a un valor M_f. Ver Tabla 5.11.

Cuadro 5.11: Enumeración de los valores de las variables ordinales

Posible valor	**Valor asociado**
Malo	1
Aceptable	2
Regular	3
Bueno	4
Excelente	5

Posible valor	**Valor asociado**
Bronce	1
Plata	2
Oro	3

Se aplica el paso 2 (a), Reemplazar el valor de cada variable ordinal (r_{if}) en la matriz de datos por su respectivo valor en la escala, ver Tabla 5.12.

Cuadro 5.12: Transformado la matriz de datos de las variables ordinales

$CódigoEstudiante$	$Atributo_1$	$Atributo_2$	$Atributo_3$	$Atributo_4$
1	5	2	2	3
2	3	4	3	3
3	1	1	1	1
4	1	4	1	2
5	4	3	4	3
6	1	3	3	2

Se aplica el paso 2 (b), aplicarle la formula 5.10, ver Tabla 5.13.

Cuadro 5.13: Transformado la matriz de datos de las variables ordinales

$CódigoEstudiante$	$Atributo_1$	$Atributo_2$	$Atributo_3$	$Atributo_4$
1	1,00	0,25	0,25	1,00
2	0,50	0,75	0,50	1,00
3	0,00	0,00	0,00	0,00
4	0,00	0,75	0,00	0,50
5	0,75	0,50	0,75	1,00
6	0,00	0,50	0,50	0,50

Finalmente se aplica el paso 3, trabajarla como una variable definida en un intervalo, usando por ejemplo uno de los casos particulares de la métrica de ***Minkowski*** 5.1 en la página 40.

5.1.1.5. Variables de tipos mixtos

Ya se han estudiado la disimilaridad entre objetos descritos por variables de diferentes tipos. Sin embargo en la práctica, los objetos son descritos por un conjunto de variables no necesariamente del mismo tipo, sino de tipos mixtos.

La disimilaridad entre dos objetos i y j puede ser calculada con base a la ecuación 5.11.

$$d(i,j) = \frac{\sum_{f=1}^{p} \sigma_{ij}^{(f)} d_{ij}^{(f)}}{\sum_{f=1}^{p} \sigma_{ij}^{(f)}} \tag{5.11}$$

En donde:

- p: representa el número de variables.
- $\sigma_{ij}^{(f)}$: Toma el valor de cero (0) o de uno (1) según uno de los siguientes casos:
  - $\sigma_{ij}^{(f)} = 0$ si:
    - x_{if} o x_{jf} no coinciden en el tipo de función que mide la variable f sobre le $objeto_i$ y el $objeto_j$.
    - $x_{if} = x_{jf} = 0$, pero la variable f es una variable binaria asimétrica.
  - $\sigma_{ij}^{(f)} = 1$, en cualquier otro caso
- $d_{ij}^{(f)}$: Toma el valor según uno de los siguientes casos y dependiendo del tipo de variable f:

Cuadro 5.14: Ejemplo de variables de tipos mixtos

$CódigoEstudiante$	Intervalo		Binaria		Nominal		Ordinal	
	$Atrib_1$	$Atrib_2$	$Atrib_3$	$Atrib_4$	$Atrib_5$	$Atrib_6$	$Atrib_7$	$Atrib_8$
1	2,0	5,0	0	0	a	u	Excelente	Aceptable
2	4,0	2,0	1	0	u	u	Regular	Bueno
3	2,0	1,0	1	1	e	i	Malo	Malo
4	4,0	0,0	1	0	i	u	Malo	Bueno
5	1,0	3,0	1	0	i	e	Bueno	Regular
6	2,0	2,0	0	1	i	o	Malo	Regular

- **Variable basada en intervalos**: (normalizar la variable entre [0...1])

$$d_{ij}^{(f)} = \frac{|x_{if} - x_{jf}|}{\max_h x_{hf} - \min_{\mathrm{h}} x_{hf}}$$

  En donde:

  - h: se mueve entre todas las variables basadas en intervalos.

- **Binaria o Nominal**: Va a tomar el valor según:
  - $d_{ij}^{(f)} = 0$ si $x_{if} = x_{jf}$
  - $d_{ij}^{(f)} = 1$ en otro caso

- **Ordinal**:
  - Aplicar el procedimiento para variables ordinales, aplicando la ecuación 5.10.

El ejemplo que aparece en la Tabla 5.14, permite ilustrar la forma de calcular la disimilaridad (ver Tabla 5.15) .

Cuadro 5.15: Distancias sobre el ejemplo de variables de tipos mixtos

	1	2	3	4	5	6
1	0,00	–	–	–	–	–
2	0,62	0,00	–	–	–	–
3	0,86	0,70	0,00	–	–	–
4	0,65	0,29	0,64	0,00	–	–
5	0,59	0,53	0,77	0,47	0,00	–
6	0,72	0,79	0,64	0,61	0,68	0,00

5.1.1.6. Variables tipo vector

En algunas aplicaciones en las que se quiere medir la similaridad, por ejemplo del contenido de dos archivos, o dos secuencias de ADN; es necesario abandonar las métricas tradicionales de distancia, para introducir una función de similaridad no métrica.

Existen muchas formas de definir esta función de similaridad, pero la más común es la medida del coseno (ver ecuación 5.14)

$$s(x, y) = \frac{x^t.y}{\|x\| \, \|y\|} \tag{5.12}$$

En donde:

- x^t: es la transpuesta del vector x.
- $\|x\|$: es la norma euclidiana del vector x, lo cual equivale a la ecuación 5.13.

$$\|x\| = \sqrt{x_1^2 + x_2^2 + x_3^2 + ... + x_p^2} \tag{5.13}$$

- $s(x, y)$: es el coseno del ángulo entre el vector x y el vector y.

  Este valor no es afectado por rotaciones y escalamientos que sufran los vectores, sin embargo si es afectado por traslaciones y transformaciones lineales en general.

El ejemplo que aparece en la Tabla 5.6 en la página 44, permite ilustrar la forma de calcular la disimilaridad (ver Tabla 5.16) .

Cuadro 5.16: Distacias sobre el ejemplo de variables tipo vector

Disimilaridad

	1	2	3	4	5	6
1	0,00	–	–	–	–	–
2	1,00	0,00	–	–	–	–
3	1,00	0,42	0,00	–	–	–
4	1,00	0,29	0,18	0,00	–	–
5	0,42	0,42	0,33	0,18	0,00	–
6	1,00	1,00	0,18	0,50	0,59	0,00

Similaridad

	1	2	3	4	5	6
1	0,00	–	–	–	–	–
2	0,00	0,00	–	–	–	–
3	0,00	0,58	0,00	–	–	–
4	0,00	0,71	0,82	0,00	–	–
5	0,58	0,58	0,67	0,82	0,00	–
6	0,00	0,00	0,82	0,50	0,41	0,00

Una variante de la función de similaridad es la llamada **coeficiente de Tanimoto** o **distancia Tanimoto**, ver ecuación 5.14.

$$s(x, y) = \frac{x^t.y}{x^t.x + y^t.y - x^t.y} \tag{5.14}$$

Si se utiliza esta distancia, el resultado para este ejemplo en particular concuerda con la similaridad obtenida con variables binarias asimétricas (ver 5.7 en la página 44).

5.1.2. Métodos de clasificación (Clustering)

Como se puede observar en [1], [2] y especialmente en [15], existen muchos algoritmos de clasificación y hablar de todos ellos sería una tarea que sobrepasa el alcance de este capítulo. Sin embargo, es posible describir las categorías básicas en las cuales se pueden agrupar estos algoritmos, dependiendo de la forma que emplee para realizar la clasificación. Una salvedad a realizar, es que estas categorías pueden solaparse parcialmente entre sí.

Las categorías a las que se hacen referencia son: particionamiento, jerárquicos, basados en densidad, basados en rejilla, basados en modelos, datos de altas dimensiones, basado en restricciones. Existen otras clasificaciones como por ejemplo detección de fronteras (basados en: distribución estadística, distancia, densidad, desviación), las cuales no se discutirán en este documento.

5.1.2.1. Métodos de particionamiento

Dados n objetos, estos se clasifican en k clases o aglomeraciones, en donde $k \leq n$. Para ser válida la clasificación se deben satisfacer las siguientes condiciones:

- Cada clase debe contener al menos un objeto.
- Cada objeto se debe encontrar exactamente en una clase.

La idea inicial consiste en crear una partición inicial y usando una técnica de relocalización interactiva de objetos, estos se mueven de una clase a otra. Una buena partición sería hacer que todos los objetos de una misma clase estén muy cercanos, mientras que los objetos de clases diferentes, estén muy separados. Existen al menos dos métodos heurísticos ampliamente utilizados: *k-means* y *k-medoids*. La diferencia entre ellos es lo que representa cada clase, en el primero, lo hace el valor medio de los objetos de la clase; mientras que en el segundo lo hace el objeto más cercano al centro de la clase. Estos métodos heurísticos, son buenos cuando el volumen de objetos a clasificar es pequeño o mediano, y ellos se encuentran agrupados en clases con forma esférica.

Un ejemplo de un algoritmo de este método es:

- **CLARA** (***C****lustering* ***LAR****ge* ***A****pplications*)

5.1.2.2. Métodos jerárquicos

Existen dos aproximaciones a estos métodos de clasificación de objetos:

- Aproximación de abajo hacia arriba, conocida como ***AGNES*** (***AG****lomerative* ***NES****ting*): Consiste en tener cada objeto en una clase diferente luego, se van fusionando de forma sucesiva las clases que se encuentran cercanas, produciendo un nuevo nivel en la jerarquía. Esto se hace hasta obtener una sola clase: el nodo raíz de la jerarquía; o hasta cumplir una condición de terminación predeterminada.

- Aproximación de arriba hacia abajo, conocida como ***DIANA*** (***DI****visive* ***ANA****lysis*): Consiste en tener todos los objetos en una única clase y luego, iteración por iteración, se va dividendo esta clase en subclases más pequeñas, produciendo un nuevo nivel en la jerarquía. Esto se hace hasta obtener cada objeto en una clase independiente: hojas de la jerarquía; o hasta cumplir una condición de terminación.

Estos métodos pueden ser considerados rígidos en el sentido que una vez hecha una fusión o una división, esta operación no puede ser deshecha. Esta característica tiene dos consideraciones: disminuye los costos computacionales al manejar el número combinatorio de las posibles selecciones; al tiempo que no se pueden corregir las decisiones erróneas en las fusiones o en las divisiones. La estructura de árbol que se produce con estos métodos se llama dendrograma.

Ejemplos de algoritmos de este método son:

- **BIRCH** (***B****alanced* ***I****terative* ***R****educing and* ***C****lustering* ***U****sing* ***H****ierarchies*)

- **ROCK** (***RO****bust Clustering using lin****Ks***)

- ***Chameleon***

5.1.2.3. Métodos basados en densidad

Como se mencionó en la sección 5.1, un requerimiento sobre los algoritmos de aglomeración es el poder descubrir agrupaciones en formas arbitrarias[1] y es precisamente esta característica que lo diferencia del resto de métodos. A grandes rasgos, la idea consiste en ir aumentando la densidad, número de objetos en la vecindad de una agrupación, hasta exceder un umbral determinado.

Ejemplos de algoritmos de este método son:

[1] la mayoría de los algoritmos de clasificación se basan en medidas de distancia Euclidiana o Manhanttan, las cuales producen sólo formas esféricas.

- **DBSCAN** (*A **D**ensity-**B**ased Clustering Method Based on Connected Regions with Sufficiently High Density*)
- **OPTICS** (***O**rdering **P**oin**T** to **I**dentify the **C**lustering **S**tructure*)
- **DENCLUE** (*Clustering Based on Density Distribution Fuctions*)

5.1.2.4. Métodos basados en rejilla

Los algoritmos basados en estos métodos mapean el espacio de objetos a un número finito de celdas que forman una estructura de rejilla. Entre las principales ventajas se encuentran la velocidad de procesamiento y la independencia del número de objetos. Por supuesto existe una dependencia sobre el número de celdas.

Ejemplos de algoritmos de este método son:

- **STING** (***ST**atistical **IN**formation **G**rid*)
- **WaveCluster** (***Cluster**ing Using **Wave**let Transformation*)

5.1.2.5. Métodos basados en modelos

Los algoritmos basados en estos métodos intentan ajustar los datos a un modelo dado. Pueden localizar aglomeraciones construyendo funciones de densidad que reflejen la distribución espacial de los objetos.

Ejemplos de algoritmos de este método son:

- **EM** (***E**xpectation **M**aximization*)
- **COBWEB** (*Incremental Conceptual Clustering*)
- **SOM** (***S**elf-**O**rganizaing Feature **M**ap*)

5.1.2.6. Métodos de datos de altas dimensiones

Los algoritmos que basan su funcionamiento en estos métodos son de mucha importancia hoy en día, ya que muchas aplicaciones requieren del análisis de objetos que contienen un gran número de características (dimensiones). Este es el caso por ejemplo de la clasificación de documentos por medio de palabras clave.

Ejemplos de algoritmos de este método son:

- **CLIQUE** (***Cluster****ing* ***I****n* ***QUE****st*)
- **PROCLUS** (***PRO****jected* ***CLUS****tering*)
- **PCluster** *(**P**attern **Cluster**)*

5.1.2.7. Métodos basados en restricciones

Los algoritmos que utilizan estos métodos, se caracterizan en que incorporan restricciones del usuario o describen propiedades que deben satisfacer las aglomeraciones resultantes o los objetos individuales. Como por ejemplo el número esperado de aglomeraciones, el mínimo o máximo tamaño de ellas, el valor de uno o varios atributos de los objetos, entre otras.

Las restricciones se pueden clasificar en tres categorías a saber: Restricciones sobre objetos individuales, restricciones sobre la selección de parámetros de aglomeración y restricciones sobre la distancia o función de similaridad.

Ejemplos de algoritmos de este método son:

- *Clustering with Obstacle Object*
- *User-Constrained Cluster Analysis*
- *Semi-Supervised Cluster Analysis*

5.2. Objetos de aprendizaje

En el contexto de esta tesis, se debe entender por objeto de aprendizaje lo expresado por David Wiley [44]: “Un objeto de aprendizaje es un recurso digital que puede ser reutilizado para ayudar en el aprendizaje”. Esta definición muy general se puede complementar con lo expresado por otros autores como Jhon Stephenson [40], quien expresa que un objeto de aprendizaje es un elemento pequeño autocontenido, que se puede almacenar y retirar separadamente a demanda de varias fuentes, según los criterios de selección aplicados. Pero el mismo autor dice que para que esto pueda darse es necesario contar con estándares por los que se pueda identificar cualquier objeto de aprendizaje y faciliten su reutilización. Y aclara que este proceso se denomina metaetiquetado electrónico.

Por supuesto que estas definiciones cubren un abanico muy grande de elementos, es por eso que el mismo autor David Wiley define una taxonomía de tipos de objetos de aprendizaje:

- **Objeto de aprendizaje fundamental**: Es un objeto digital que sólo contiene información.

- **Objeto de aprendizaje combinado cerrado**: Es un conjunto de recursos digitales que constituyen una unidad de información.
- **Objeto de aprendizaje combinado abierto**: Es similar a un objeto de aprendizaje combinado cerrado, pero con la característica que admite nuevos elementos.
- **Presentación generativa de objetos de aprendizaje**: Es similar a un objeto de aprendizaje combinado cerrado, pero incluyen actividades evaluativas.
- **Objeto de aprendizaje generativo instruccional**: Es similar a un objeto de aprendizaje combinado cerrado, pero incluye prácticas sobre un procedimiento en particular.

5.2.1. Características de un objeto de aprendizaje

Según Josep M. Boneu[3], "para la reutilización e interoperabilidad de contenidos en diferentes plataformas, debe cumplirse una doble premisa: por un lado los cursos deben seguir un estándar y por otro lado las plataformas deben soportar dicho estándar, con lo que se facilita el uso de cursos realizados por la propia organización y por terceros".

Además dice que existen estándares como:

- El estándar LOM[2] de IEEE LTSC[3], ampliamente aceptado, permite describir el contenido de un objeto de aprendizaje a través de metadatos.
- La AICC[4] fue el primer organismo creado para desarrollar un conjunto de normas que permitieran el intercambio de cursos CBT.
- La descripción de itinerarios formativos es posible mediante las estructuras de descripción de recursos RDF[5] o bien a través de IMS LD[6], el cual está basado en EML[7].
- Y finalmente el estándar SCORM[8] de ADL[9], ampliamente aceptado y utilizado, que permite la integración de los diferentes estándares en un solo elemento empaquetado:
    - El metadatos (Estándar IEEE - LOM 1481.12.1-2002).
    - La estructura de contenidos (Estándar AICC).
    - El empaquetamiento de contenidos (Estándar IMS LD).
    - El secuenciamiento de información (Estándar IMS LD).

[2]LOM: Learning Object Metadata.
[3]IEEE LTSC: Institute for Electrical and Electronic Engineers - Learning Technology Standards Committee.
[4]AICC: Aviation Industry CBT Committee.
[5]RDF: Resource Description Framework.
[6]IMS LD: IMS learning design.
[7]EML: Educational Modelling Language.
[8]SCORM: Shareable Content Object Reference Model.
[9]ADL: Advanced Distribute Learning.

5.2.2. Metadatos de un objeto de aprendizaje

Como ya se ha mencionado el estándar IEEE - LOM 1481.12.1-2002 [18] permite describir el contenido de un objeto de aprendizaje a través de metadatos.

Antes de iniciar un recorrido rápido por el estándar LOM, es importante indicar que un objeto de aprendizaje en este estándar es definido como "cualquier entidad, digital o no, susceptible de ser usada en aprendizaje, educación o formación.". Es de resaltar en la definición, que es posible describir el contenido de un objeto de aprendizaje ya sea digital o no.

El estándar LOM permite describir las características relevantes de un objeto de aprendizaje. Dichas características se pueden agrupar en las categorías: General, Ciclo de vida, Meta-metadatos, Técnica, Uso educativo, Derechos, Relación, Anotación y Clasificación.

Como lo indica el mismo estándar, su propósito es facilitar la búsqueda, evaluación, adquisición y uso de los objetos de aprendizaje, por ejemplo, por estudiantes, profesores o procesos automáticos de software. El estándar también facilita el intercambio y uso compartido de objetos de aprendizaje, permitiendo el desarrollo de catálogos e inventarios.

5.2.2.1. Estructura de un metadato

Tomando del estándar, el esquema base LOM 1.0 (cláusula 6) consiste en nueve de estas categorías.

- **General**: agrupa la información general que describe un objeto de aprendizaje de manera global. Por ejemplo: identificador (catálogo, entrada), título, idioma, descripción, palabra clave, ámbito, estructura y nivel de agregación.
- **Ciclo de Vida**: agrupa las características relacionadas con la historia y el estado actual del objeto de aprendizaje, y aquellas que le han afectado durante su evolución. Por ejemplo: versión, estado y contribución (tipo, entidad, fecha).
- **Meta-Metadatos**: agrupa la información sobre la propia instancia de Metadatos, (en lugar del objeto de aprendizaje). Por ejemplo: identificador (catálogo, entrada), contribución (tipo, entidad, fecha), esquema de metadatos e idioma.
- **Técnica**: agrupa los requerimientos y características técnicas del objeto de aprendizaje. Por ejemplo: formato, tamaño, localización, requisitos, pautas de instalación, otros requisitos de plataforma y duración.
- **Uso Educativo**: agrupa las características educativas y pedagógicas del objeto de aprendizaje. Por ejemplo: tipo de interactividad, tipo de recurso educativo, nivel de interactividad, densidad semántica, destinatario, contexto, rango típico de edad, dificultad, tiempo típico de aprendizaje, descripción e idioma.

- **Derechos**: agrupa los derechos de propiedad intelectual y las condiciones para el uso del objeto de aprendizaje. Por ejemplo: costo, derechos de autor y otras restricciones.
- **Relación**: agrupa las características que definen la relación entre este objeto de aprendizaje y otros objetos relacionados. Por ejemplo: tipo, recurso (identificador, catálogo, entrada) y descripción.
- **Anotación**: permite incluir comentarios sobre el uso educativo del objeto e información sobre cuándo y por quién fueron creados dichos comentarios. Por ejemplo: entidad, fecha y descripción.
- **Clasificación**: describe este objeto de aprendizaje en relación a un determinado sistema de clasificación, permitiéndole a un usuario final clasificar un objeto de aprendizaje de acuerdo con una estructura de clasificación arbitraria (esta categoría se proporciona como un mecanismo de extensión).

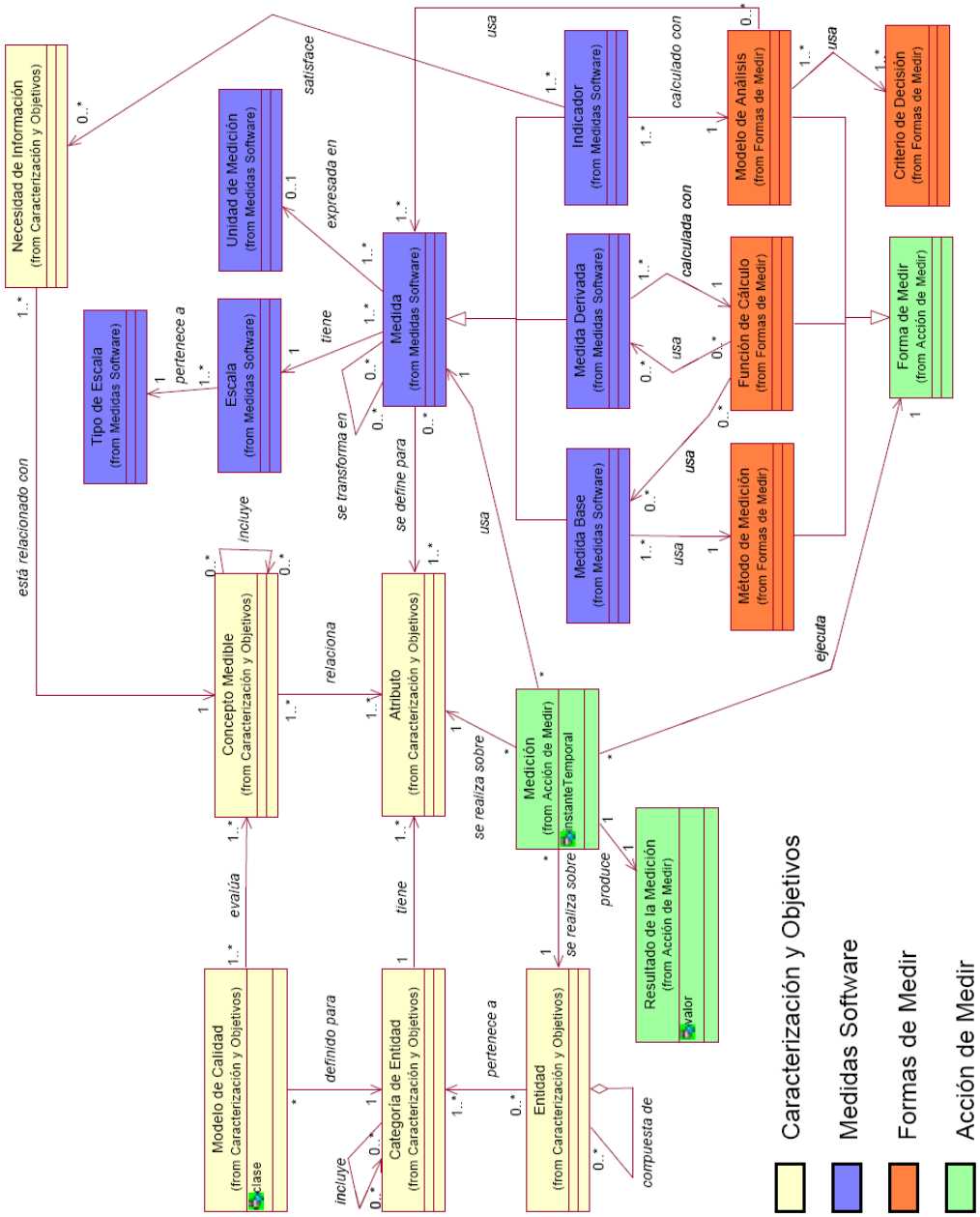
5.3. Proceso de medición

Esta sección esta basada principalmente en el "Informe Técnico UCLM-TSI-001 sobre Medición del Software Ontología y Metamodelo"[13], pues en sus palabras el objetivo principal de ellos es "dar soporte a la definición de modelos concretos de medición, facilitando así su gestión y proporcionar un marco adecuado para el proceso de medición" y motivados principalmente por el hecho de que no hay todavía consenso sobre los conceptos y terminologías utilizadas en este campo.

Aunque es cierto que la Figura 5.2 muestra la ontología de la medición del software, estos conceptos pueden perfectamente ser extendidos a la medición de objetos de aprendizaje, y así poder tener un marco adecuado para el proceso de medición de ellos. Es apropiado aclarar que esta ontología es el resultado de analizar diferentes fuentes como:

- Estándar IEEE 610.12-1992 (*Standard Glossary of Software Engineering Terminology*).
- Estándar IEEE 1061-1998 (*IEEE Standard for a Software Quality Metrics Methodology*).
- Estándar ISO/IEC 14598:2001 series (*Software engineering - Product evaluation*).
- Estándar ISO VIM (*International Vocabulary of Basic and General Terms in Metrology*).
- Estándar ISO/IEC 15939 (*Software engineering - Software measurement process*).

Figura 5.2: Diagrama de la Ontología de la Medición del Software



Fuente: *Informe Técnico UCLM-TSI-001 sobre Medición del Software Ontología y Metamodelo* [13]

- Ontología del modelado de la calidad en la empresa de Henry Kim.
- Método para la especificación de modelos de conjuntos de datos del software de Kitchenham *et al*.
- Método GQM/MEDEA para la definición de medidas de atributos de productos en la ingeniería del software de Briand *et al*.

Una vez hecha esta aclaración es bueno hablar un poco de esta ontología. Los autores M. Ferreira, F. García, F. Ruiz *et al*[13], dividieron la ontología entre cuatro sub-ontologías, las cuales en sus palabras son definidas como:

- **Caracterización y Objetivos de la Medición Software**, que incluye los conceptos necesarios para establecer el ámbito y los objetivos del proceso de medición software. El principal objetivo del proceso de medición software es satisfacer ciertas `necesidades de información` identificando las `entidades` (que pertenecen a `categoría de entidad`) y los `atributos` de estas `entidades` (que son el enfoque del proceso de medición). `Atributos y necesidades de información` están relacionados por medio de `conceptos medibles` (que pertenecen a `modelo de calidad`).
- **Medidas Software**, que trata de establecer y aclarar los elementos clave en la definición de una `medida software`. Una `medida` relaciona un `método de medición` definido y una `escala` de medición (que pertenece a `tipo de escala`). La mayoría de las `medidas` pueden ser expresadas en una `unidad de medición`, y pueden ser definidas para más de un `atributo` (Medidas nominales son ejemplos de medidas que no pueden ser expresadas en unidades de medición). Las `medidas` pueden ser de 3 tipos: `medidas base`, `medidas derivadas` e `indicadores`.
- **Formas de Medir**. Esta sub-ontología introduce el concepto de `forma de medir` para generalizar los diferentes "métodos" usados por los tres tipos de `medidas` para obtener sus respectivos `resultados de medición`. Una `medida base` aplica un `método de medición`. Una `medida derivada` usa una `función de cálculo` (que se apoya en otra medida base y/o derivada). Por fin, un `indicador` utiliza un `modelo de análisis` (basado en un `criterio de decisión`) para obtener un resultado de medición que satisfaga a una `necesidad de` información.
- **Medición**. Establece la terminología relacionada con el acto de medir el software. Una `medición` (que es una acción) es un conjunto de operaciones cuyo objetivo es determinar el valor de un `resultado de medición` para un dado `atributo` de una `entidad`, utilizando una `forma de medir`. Los `resultados de la medición` se obtienen a través de la acción de llevar a cabo una `medición`.

5.3.1. Modelo de calidad[10]

Usando la información de la Figura 5.2 y las definiciones dadas por los autores M. Ferreira, F. García, F. Ruiz *et al* en [13], se puede decir que:

- Un `modelo de calidad` evalúa uno o muchos `conceptos medibles` y está definido para una `categoría de entidad`.

  "Un conjunto de conceptos medibles y relaciones entre ellos que proporciona la base para especificar requisitos de calidad y evaluar la calidad de las entidades de una determinada categoría de entidad".

- Un `concepto medible` relaciona uno o muchos `atributos` y puede incluir otros `conceptos medibles`.

  "Relación abstracta entre atributos y necesidades de información".

- Una `necesidad de información` se relaciona siempre con un `concepto` medible.

  "Información necesaria para gestionar un proyecto (sus objetivos, hitos, riesgos y problemas)".

- Una `categoría de entidad` puede incluir otras `categorías de entidad` y tiene asociado uno o muchos `atributos`.

  "Una colección de entidades caracterizadas por satisfacer un cierto predicado común".

- Una `entidad` pertenece a una `categoría de entidad` y puede estar compuesta por otras `entidades`.

  "Un objeto que va a ser caracterizado mediante una medición de sus atributos".

- Un `atributo` es:

  "Una propiedad mensurable, física o abstracta, que comparten todas las entidades de una categoría de entidad".

5.3.2. Medida[11]

Usando la información de la Figura 5.2 y las definiciones dadas por los autores M. Ferreira, F. García, F. Ruiz *et al* en [13], se puede decir que:

[10] Llamado en el informe técnico como "Caracterización y Objetos".
[11] Llamado en el informe técnico como "Medidas de software".

- Una `medida` se puede transformar en otras `medidas`, tiene asociada una `escala` de medición, se expresa en una `unidad de medida` y se define para uno o muchos `atributos` en particular. Una medida puede ser: una `medida base`, una `medida derivada`, un `indicador`.

  "La forma de medir (método de medición, función de cálculo o modelo de análisis) y la escala de medición.".

- Una `escala` pertenece a una `tipo de escala` en particular.

  "Un conjunto de valores con propiedades definidas".

- Una `unidad de medida` es:

  "Una cantidad particular, definida y adoptada por convención, con la que se puede comparar otras cantidades de la misma clase para expresar sus magnitudes respecto a esa cantidad particular".

- Un `tipo de escala`:

  "Indica la naturaleza de la relación entre los valores de la escala (por ejemplo: basada en intervalos, nominales, ordinales)".

- Una `medida base` es una `medida` que usa un `método de medición`.

  "Es una medida de un atributo que no depende de ninguna otra medida, y cuya forma de medir es un método de medición".

- Una `medida derivada` es una `medida` que es calculada con una `función de cálculo`.

  "Es una medida que es derivada de otra medida base o derivada, utilizando una función de cálculo como forma de medir".

- Un `indicador` es una `medida` que es calculada con un `modelo de análisis` y puede satisfacer muchas `necesidades de información`.

  "Es una medida que es derivada de otras medidas utilizando un modelo de análisis como forma de medir".

5.3.3. Formas de medir

Usando la información de la Figura 5.2 y las definiciones dadas por los autores M. Ferreira, F. García, F. Ruiz *et al* en [13], se puede decir que:

- Una `forma de medir` puede ser mediante un `método de medición`, una `función de cálculo` o un `modelo de análisis`.

  "Secuencia de operaciones cuyo objeto es determinar el valor del resultado de la medición. Una forma de medir puede ser un método de medición, función de cálculo o modelo de análisis.".

- Un `método de medición` es:

  "La forma de medir una medida base. Secuencia lógica de operaciones, descritas de forma genérica, usadas para realizar mediciones de un atributo respecto de una escala específica.".

- Una `función de cálculo` puede hacer uso de `medidas base` o de `medidas derivadas`.

  "La forma de medir una medida derivada. Algoritmo o cálculo realizado para combinar dos o más medidas base y/o derivadas.".

- Un `modelo de análisis` usa una o muchas `medidas` y usa un o muchos `criterios de decisión`.

  "La forma de medir un indicador. Algoritmo o cálculo realizado para combinar una o más medidas (base, derivadas o indicadores) con criterios de decisión asociados".

- Un `criterio de decisión` es:

  "Valores umbral, objetivos, o patrones, usados para determinar la necesidad de una acción o investigación posterior, o para describir el nivel de confianza de un resultado dado".

5.3.4. Acción de medir[12]

Usando la información de la Figura 5.2 y las definiciones dadas por los autores M. Ferreira, F. García, F. Ruiz *et al* en [13], se puede decir que:

- Una `medición` es la acción de ejecutar una `forma de medir`, la cual se realiza sobre un `atributo` en particular de una `entidad`, usando una `medida` y produciendo un `resultado de medición`.

  Una medición es un "Conjunto de operaciones que permite obtener el valor del resultado de la medición para un atributo de una entidad, usando una forma de medir".

[12] Llamado en el informe técnico como "Medición".

- Un resultado de medición es:

  "Una categoría o número asignado a un atributo de una entidad como resultado de una medición".

Teniendo este proceso de medición como base, en el capítulo siguiente, específicamente en la Sección 6.3 en la página 71 y en la Sección 6.4 en la página 76, se va a desarrollar un ejemplo particular, con el único objetivo de mostrar la aplicación del mecanismo formal de clasificación automática, desarrollado como objetivo general de esta tesis.

CAPÍTULO 6

MECANISMO FORMAL DE CLASIFICACIÓN

6.1. Espacio de medida

Antes de comenzar a definir formalmente el espacio de medida, es necesario hacer tres definiciones fundamentales basadas en el libro de Bruce K. Driver "Analysis Tools with Applications"[10].

Definición 6.1 Sea X un conjunto. Una sigma álgebra de subconjuntos de X es una clase Υ que cumple las siguientes propiedades:

- $X \in \Upsilon$,
- Si $A \in \Upsilon$, entonces $A^c \in \Upsilon$,
- Si $A_1, A_2 \ldots \in \Upsilon$, entonces $\bigcup_{n=1}^{\infty} A_n \in \Upsilon$

Definición 6.2 Una función de medida μ definida sobre Υ es una aplicación de $\mu : \Upsilon \to [0, +\infty)$ que cumple las siguientes propiedades:

- $\mu(\emptyset) = 0$,
- Si $A_1, A_2, \ldots$ son disyuntos entonces$\mu(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n), \quad \forall A_i \in \Upsilon.$

Definición 6.3 Sea X un conjunto. Sea σ una sigma álgebra de subconjuntos de X. Sea μ una función de medida en σ. Entonces se define un espacio de medida como la terna $\langle X, \sigma, \mu \rangle$.

Una vez hechas estas definiciones fundamentales, se procede a definir, como aporte de esta tesis, el espacio de medida para objetos de aprendizaje de la siguiente forma:

Definición 6.4 Un objeto de aprendizaje O se define como la terna $\langle C, S, P \rangle$ en donde C es el contenido del objeto de aprendizaje, S es el conjunto de los valores que toman los atributos definidos en el metadato asociado al objeto y P es el conjunto de todos los pares ordenados (i, v_i) en donde i es el código que identifica al indicador y v_i es el valor asociado a él.

Definición 6.5 Un repositorio de objetos de aprendizaje R se define como $R = \{O_1, O_2, ..., O_m\}$, $m \geq 0$. Si $m = 0$, se dice que el repositorio se encuentra vacio.

A partir de aquí considere X de la Definición 6.1 como el repositorio R y el sigma álgebra como partes de R. Como ya es sabido, partes de R es la mayor sigma álgebra que puede ser construida a partir de este conjunto.

Definición 6.6 Sea T el conjunto de categorias de entidad para objetos de aprendizaje. Entonces $\forall x \in R$, x debe tener asociada una y sólo una categoría $t \in T$.

Algunas categorías de entidad para objetos de aprendizaje pueden ser: presentaciones, videos, unidades, cursos. Hay que recordar que "categoría de entidad" es tomado de la Figura 5.2 en la página 58 de la Sección 5.3 del marco teórico.

Definición 6.7 Una categoría de entidad t para un objeto de aprendizaje, se define como la terna $\langle n, m, V \rangle$ en donde n es el nombre la categoria, $m = \|\{x \in R : x \text{ esté asociada la categoría } t\}\|$, y V es el conjunto de todas las ternas (i, d, v) en donde i es el identificador del indicador, d representa la descripción del identificador y v es la suma de este indicador $\forall x \in R$ y que x esté asociado a la categoría t.

Para este mecanismo de clasificación en particular, se consideran sólo los indicadores que puedan ser valorados numéricamente (Ver primer objetivo específico de la página 30). Sin embargo, esto no debe ser visto como una debilidad, por el contrario, debe ser visto como una oportunidad para futuras investigaciones. La idea, sería tomar la información de la Sección 5.1.1 en la página 39 y crear una nueva función de medida, que por supuesto hay que demostrar matemáticamente, tal y como se hará a continuación con la función propuesta.

Definición 6.8 Sea la función $h(x) = \left\|V^{(x)}\right\|$ definida como $h : X \to \mathbb{Z}^+$, en donde $\left\|V^{(x)}\right\|$ representa la cardinalidad del conjunto V de la categoría de entidad asociada al objeto x.

En otras palabras, $h(x)$ indica el número de indicadores asociados a la categoría de clase a la que pertenece el objeto de aprendizaje x.

Definición 6.9 Sea la función g, definida como $g : X \to [0, +\infty)$, en donde $g(x)$ representa la cantidad de indicadores del objeto de aprendizaje x, cuyo valor en P, se encuentra por encima del promedio (inclusive) de este indicador en la categoría de entidad a la que corresponde el objeto de aprendizaje x.

Para hacer esto, se debe consultar V y usando el valor m, calcula el valor del promedio, en el t asociado al objeto de aprendizaje x.

Teorema 6.1 Se define la función de medida μ (porcentaje de aceptación) sobre un elemento de Υ como una aplicación de $\mu : \Upsilon \rightarrow [0, +\infty)$ así: $\mu(Y) = \begin{cases} \sum_{n=1}^{\|Y\|} \frac{g(y_n)}{h(y_n)} & \text{si } \|Y\| \geq 1 \\ 0 & \text{si } \|Y\| = 0 \end{cases}$,

$\forall Y \in \Upsilon$ y $\forall y_i \in Y$.

Demostración por inducción

- $\mu(\emptyset) = 0$, se cumple por definición de la función μ cuando $\|Y\| = 0$,
- Si $A_1, A_2, \ldots$ son disyuntos, $\mu(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n), \quad \forall A_i \in \Upsilon$
  - Mostrar para $\|Y\| = 1$, $\mu(A_1) = \frac{g(y_1)}{h(y_1)}, \quad A_1 \in Y$.

    $\mu(\bigcup_{n=1}^{1} A_n) = \mu(A_1) = \frac{g(y_1)}{h(y_1)} = \sum_{n=1}^{1} \mu(A_n)$
  - Suponer para $\|Y\| = m$, $\mu(\bigcup_{n=1}^{m} A_n) = \sum_{n=1}^{m} \mu(A_n), \quad \forall A_i \in Y$ y demostrar para $\|Y\| = m + 1$.

    Por hipótesis los elementos $A_1, A_2, \ldots$ son disyuntos, entonces:

    $\mu(\bigcup_{n=1}^{m+1} A_n) = \mu(\bigcup_{n=1}^{m} A_n \cup A_{m+1}) = \mu(\bigcup_{n=1}^{m} A_n) + \mu(A_{m+1}) =$

    $\sum_{n=1}^{m} \mu(A_n) + \mu(A_{m+1}) = \sum_{n=1}^{m} \frac{g(y_n)}{h(y_n)} + \frac{g(y_{m+1})}{h(y_{m+1})} =$

    $\sum_{n=1}^{m+1} \frac{g(y_n)}{h(y_n)} = \sum_{n=1}^{m+1} \mu(A_n)$

Teorema 6.2 Se define la función de disimilaridad d sobre un par de elementos de Υ como una aplicación de $d : X \times X \rightarrow [0, +\infty)$ así: $d(x, y) = |\mu(x) - \mu(y)|, \forall x, y \in \Upsilon$.

Pero esta función debe satisfacer las siguientes propiedades, definidas previamente en la página 39, $\forall x, y, z \in \Upsilon$.

- $d(x, y)$, es un número positivo o cero.

  $d(x, y) = |\mu(x) - \mu(y)|$y por definición del valor absoluto, la propiedad se satisface.

- $d(x,y) = d(y,x)$

  $d(x,y) = |\mu(x) - \mu(y)|$, por conmutatividad de la resta en el valor absoluto

  $= |\mu(y) - \mu(x)| = d(y,x)$

- $d(x,x) = 0$,

  $d(x,x) = |\mu(x) - \mu(x)| = 0$.

- $d(x,y) \ \leq \ d(x,z) + d(z,y)$,

  $d(x,y) = |\mu(x) - \mu(y)| = |\mu(x) - \mu(y) + \mu(z) - \mu(z)| =$

  $|(\mu(x) - \mu(z)) + (\mu(z) - \mu(y))| \ \leq \ |(\mu(x) - \mu(z))| + |(\mu(z) - \mu(y))| = d(x,z) + d(z,y)$

$\therefore d(x,y) \ \leq \ d(x,z) + d(z,y)$

En el Teorema 6.2 se aplicó una de las propiedades del valor absoluto:

$|x+y| \ \leq \ |x| + |y|$, tambien llamada desigualdad triangular, la cual se demuestra usando otras propiedades del valor absoluto:

a. $-|x| \leq x \leq |x|$,$\forall x \in \mathbb{R}$, (propiedad del valor absoluto)

b. $-|y| \leq y \leq |y|$, $\forall y \in \mathbb{R}$,

c. $(-|x|) + (-|y|) \leq x + y \leq |x| + |y|$ (sumando **a** y **b**),

d. $-(|x| + |y|) \leq x + y \leq (|x| + |y|)$, (aplicando la propiedad $-a \leq x \leq a \Longleftrightarrow |x| \leq a$)

$\therefore |x+y| \ \leq \ |x| + |y|$

6.2. Método de clasificación

6.2.1. Condiciones iniciales

- $R = \emptyset$, el repositorio debe estar totalmente vacio.
- $T \neq \emptyset$, deben estar definidas algunas categorías de entidad.
- $m \neq \emptyset$, $\forall m \in t, \forall t \in T$, toda categoría de entidad debe tener un nombre definido.
- $n = 0$, $\forall n \in t, \forall t \in T$, toda categoría de entidad por la primera condición inicial, no debe tener objetos de aprendizaje asociados.
- $V \neq \emptyset$, $\forall V \in t, \forall t \in T$, toda categoría de entidad debe tener definidos identificadores.

- $\forall a, b \in V$, $a \neq b$,$\forall V \in t, \forall t \in T$, todos los indicadores deben ser diferentes, se define la comparación de elementos de V como la comparación de sus identificadores (i).
- $v = 0$, $\forall n \in V, \forall V \in t, \forall t \in T$, el valor de todos los indicadores de las categorias de entidad debe ser cero.

6.2.2. Condiciones para adicionar un objeto de aprendizaje O

- Identifique a qué categoría de entidad t corresponde el objeto de aprendizaje O a adicionar.
- Obtenga todos los resultados de la medición de los indicadores definidos en t para O, considerando C y S, en donde $C, S \in O$ y construya apropiadamente P con los resultados, ahora complete la terna $\langle C, S, P \rangle$.
- Adicione O al repositorio R y actualice apropiadamente el componente V de t (sume los valores obtenidos en P en el punto anterior a los indicadores correspondientes).

6.2.3. Condiciones para retirar un objeto de aprendizaje O

- Identifique a qué categoría de entidad t corresponde el objeto de aprendizaje O a retirar.
- Retire O del repositorio R y actualice apropiadamente el componente V de t (restando los valores en P del objeto de aprendizaje O).

6.2.4. Clasificación de los objeto de aprendizaje O

Construya una secuencia ordenada L, tal que $\forall x \in R,\ x \in L$ y que $\forall l_i, l_j \in L$ en donde $i < j$ se debe cumplir que $\mu(l_1) \leq \mu(l_2)$.

Una vez construida la secuencia ordenada L, los peores objetos de aprendizaje pertenecientes al repositorio R serán los ubicados en las primeras posiciones de la secuencia y por tanto serán los mejores candidatos a entrar a un proceso de mejoramiento. Una vez se ha decidido mejorar un objeto de aprendizaje, éste deberá ser retirado del repositorio, mejorado y vuelto a adicionar al mismo, una vez se hayan hecho las mejoras.

Como se puede observar, este algoritmo no considera directamente el contenido del (C) objeto de aprendizaje o su metadato (S), sino el conjunto de todos los pares ordenados (i, v_i), llamado P, en donde i es el código que identifica al indicador y v_i es el valor asociado a él. Por supuesto para poder construir a P es necesario que un personal capacitado evalue a C y/o

a S. En este orden de ideas, este algoritmo puede ser clasificado como un algoritmo que utiliza un método basado en restricciones (Ver Sección 5.1.2.7 en la página 54), pues todos los objetos de aprendizaje que tengan el mismo número de indicadores por encima del promedio (inclusive) y sin importar cuales, pertenecerán a la misma categoría.

Como ya es sabido, el mejor algoritmo de ordenamiento basado en comparaciones tiene una complejidad computacional del orden de $O(n \ln n)$, como por ejemplo el *QuickSort*, sin embargo, para este caso particular, se puede utilizar un algoritmo de ordenamiento de orden $O(n)$, llamado *RadixSort*, usándolo para ordenar $\lfloor \mu(\cdot) * 10 \rfloor$, en donde 10 restringe a una precisión de una cifra decimal (una precisión más que suficiente para esta aplicación). La explicación de este algoritmo y su implementación se sale del alcance de esta tesis, pero puede ser consultado en la literatura de ciencias de la computación, como por ejemplo el libro de Thomas H. Cormen [8] "Introduction to Algorithms", especialmente en su capítulo "Sorting in Linear Time".

Las aglomeraciones se pueden obtener al identificar los objetos de aprendizaje (entidades) que tienen igual porcentaje de aceptación, o en un defecto tengan un grado de disimilaridad pequeño (parámetro del usuario).

6.2.5. Requerimiento sobre el algoritmo

Según lo visto en la Sección 5.1 en la página 38, un algoritmo de aglomeramiento debe cumplir una serie de requermientos:

- **Escalabilidad:** El algoritmo planteado no tiene problema en trabajar con un número alto de objetos, más aún si utiliza un algoritmo de ordenamiento en tiempo lineal.
- **Trabajar con diferentes tipos de datos:** Esta primera investigación restringe a indicadores que puedan ser valorados numéricamente. Investigaciones futuras deberán ampliar el abanico en los tipos de datos utilizados para valorar los indicadores, tal y como se trabajo en la Sección 5.1.1 en la página 39.
- **Descubrir agrupaciones en formas arbitrarias:** Este algoritmo utiliza la ***Métrica Manhanttan*** (o del taxista), ver la ecuación 5.2, la cual se reduce en este caso a $d(x, y) = |\mu(x) - \mu(y)|$. Este algoritmo toma los objetos de aprendizaje y los ubica en una recta que representa el intervalo $[0{,}0, 1{,}0]$, en esta recta se ubican puntos que representan porcentajes de aceptación de un objeto de aprendizaje ($\frac{g(y_n)}{h(y_n)}$), los cuales coinciden con las aglomeraciones realizadas. Como podrá visualizarse en la Sección 6.5, los objetos pueden moverse de una aglomeración a otra, cada vez que se realice una operación sobre el repositorio R, por tanto y bajo la óptica de los objetos de aprendizaje, sería más apropiado hablar de que el algoritmo descubre formas arbitrarias.

- **Mínima cantidad de parámetros de entrada:** Como ya se ha dicho, este algoritmo requiere por cada objeto de aprendizaje O, un conjunto P de todos los pares ordenados (i, v_i) en donde i es el código que identifica al indicador y v_i es el valor asociado a él.

- **Habilidad para trabajar datos con ruido:** Si un objeto está mal evaluado, esta evaluación errónea puede (y seguramente lo hará) modificar la clasificación de los demás objetos de esa categoría, pues puede alterar los valores almacenados en la componente v de la categoría t a la que pertenece el objeto de aprendizaje mal evaluado. La ventaja es que una vez detectada la inconsistencia, este objeto puede ser retirado del repositorio, restaurando los valores anteriores de v y haciendo que se restaure la clasificación original.

- **Aglomeración incremental e insensibilidad a la cantidad de objetos de entrada:** Como se pudo observar en la Sección 6.2.2 y 6.2.3, es muy sencillo insertar o retirar objetos de aprendizaje del repositorio R. Esto por supuesto no trae nuevas complicaciones a la hora de crear la secuencia ordenada L, pues lo que hay que hacer es nuevamente ordenar los objetos de repositorio usando preferiblemente un algoritmo de tiempo lineal como el ya expuesto.

- **Alta dimensionalidad:** Este algoritmo puede trabajar con objetos de aprendizaje con un número elevado de indicadores (norma de P), es más, puede trabajar con diferentes categorías de objetos de aprendizaje, cada uno con una cantidad diferente de indicadores. La dificultad entonces no estará en el algoritmo, sino en el equipo que debe evaluar cada objeto de aprendizaje para construir a P.

- **Agrupación basada en restricciones:** Este algoritmo crea aglomeraciones que satisfagan ciertas restricciones o condiciones especificadas, en este caso el porcentaje de aceptación del objeto de aprendizaje según la valoración de un conjunto de indicadores definidos para cada categoría de objetos.

- **Interpretabilidad y usabilidad:** No debe haber discusión en que los resultados de las aglomeraciones son fácilmente interpretados, comprendidos y que pueden ser usables. En este caso, los diferentes objetos (sin importar a la clase a la que pertenezcan), van a estar clasificados en aglomeraciones que representan el porcentaje de aceptación de esos objetos de aprendizaje, tal y como ya se ha explicado en reiteradas ocasiones. La usabilidad de la información radica principalmente en la detección de los mejores candidatos a entrar a un proceso de mejoramiento.

6.3. Contexto concreto

En esta sección se intentará definir un contexto para poder crear un ejemplo en la Sección 6.5. Para realizarlo se utilizará el proceso de medición estudiado en la Sección 5.3.

Existe un riesgo al definir un contexto concreto y es que el lector desvie su atención a problemas que se encuentran fuera del alcance de esta tesis (ej: la formulación apropiada de indicadores, la identificación de conceptos medibles o de atributos significativos, entre muchos otros problemas). En este orden de ideas, tener un contexto concreto permite que el lector se focalice en comprender el mecanismo de medición definido en este capítulo.

Los atributos utilizados son tomados de la página web "TECNOLOGIA EDUCATIVA - WEB PERE MARQUÈS"[23], al igual que las categorías de entidad (Videos didácticos[1], Transparencias y Diapositivas[2], Aplicaciones multimedia[3], Páginas Web[4]).

- **Modelo de calidad:**

"Modelo de calidad para objetos de aprendizaje"

- **Conceptos medibles para el** *"Modelo de calidad para objetos de aprendizaje"*, según las necesidades de información que se tengan:

"Aspectos funcionales"	*"Aspectos Técnicos"*	*"Aspectos Pedagógicos"*

- **Categorias de entidad para el** *"Modelo de calidad para objetos de aprendizaje"*

"Video didácticos"	*"Transparencias y Diapositivas"*
"Aplicaciones multimedia educativas"	*"Páginas Web Educativas"*

- **Atributos por cada categoría de entidad del** *"Modelo de calidad para objetos de aprendizaje"*

  - *"Video didácticos"*

A_{11} :	*"Eficacia (puede facilitar el logro de sus objetivos)"*
A_{12} :	*"Relevancia curricular de los objetivos que persigue"*
A_{13} :	*"Contenidos (calidad, profundidad, organización)"*
A_{14} :	*"Estructura y ritmo (guión claro, secuenciación...)"*
A_{15} :	*"Planteamiento audiovisual (interacción entre elementos)"*
A_{16} :	*"Textos, gráficos y animaciones"*
A_{17} :	*"Capacidad de motivación (atractivo, interés)"*
A_{18} :	*"Adecuación al usuario (contenidos, actividades)"*
A_{19} :	*"Planteamiento didáctico (organizadores, resumen...)"*

[1] http://www.pangea.org/peremarques/videoav2.htm
[2] http://www.pangea.org/peremarques/diapoeva.htm
[3] http://www.pangea.org/peremarques/evalua.htm
[4] http://www.pangea.org/peremarques/evalweb.htm

- *"Transparencias y Diapositivas"*

A_{21} :	*"Eficacia. El material facilita la comprensión de las ideas"*
A_{22} :	*"Contenidos. Los contenidos son significativos, correctos, actuales y de calidad"*
A_{23} :	*"Adecuación de los contenidos a los destinatarios"*

- *"Aplicaciones multimedia educativas"*

A_{31}	*"Versatilidad didáctica: modificable, niveles, ajustes, informes..."*
A_{32}	*"Documentación, guía didáctica"*
A_{33}	*"Interacción (tipo de diálogo, entrada de datos, análisis respuestas)"*
A_{34}	*"Estructura y navegación por las actividades, metáforas"*
A_{35}	*"Potencialidad de los recursos didácticos (actividades, organizadores)"*
A_{36}	*"Posibilita el trabajo cooperativo, da facilidades para este"*

- *"Páginas Web Educativas"*

A_{41}	*"Carácter multilingüe, al menos algunos apartados principales"*
A_{42}	*"Relevancia, interés de los contenidos y servicios que ofrece"*
A_{43}	*"Ejecución fiable, velocidad de acceso adecuada"*
A_{44}	*"Fomento del autoaprendizaje. iniciativa, toma decisiones"*

- **Atributos relacionados con cada concepto medible del** *"Modelo de calidad para objetos de aprendizaje"*

  - *"Aspectos Funcionales"*

A_{11}	A_{12}	A_{21}	A_{31}	A_{32}	A_{41}	A_{42}

  - *"Aspectos Técnicos"*

A_{13}	A_{14}	A_{15}	A_{16}	A_{22}	A_{33}	A_{34}	A_{43}

  - *"Aspectos Pedagógicos"*

A_{17}	A_{18}	A_{19}	A_{23}	A_{35}	A_{36}	A_{44}

- **Entidades (objetos de aprendizaje) por cada categoría de entidad**

  - *"Video didácticos"*

"Video 1"	*"Video 2"*	*"Video 3"*
"Video 4"	*"Video 5"*	*"Video 6"*
"Video 7"	*"Video 8"*	

- *"Transparencias y Diapositivas"*

"Presentación 1"	*"Presentación 2"*	*"Presentación 3"*
"Presentación 4"	*"Presentación 5"*	*"Presentación 6"*
"Presentación 7"		

- *"Aplicaciones multimedia educativas"*

"Multimedia 1"	*"Multimedia 2"*	*"Multimedia 3"*
"Multimedia 4"	*"Multimedia 5"*	*"Multimedia 6"*
"Multimedia 7"	*"Multimedia 8"*	*"Multimedia 9"*

- *"Páginas Web Educativas"*

"Web 1"	*"Web 2"*	*"Web 3"*
"Web 4"	*"Web 5"*	*"Web 6"*

Una vez definido por cada atributo su forma de medición (Método de medición, Función de cálculo, o por un modelo de análisis) se procede a su medición por parte de un equipo competente, obteniendo así unos resultados de medición por cada entidad (objeto de aprendizaje), según la categoría a la que pertenezca.

- *"Video didácticos"*

$Entidad/Atributo$	A_{11}	A_{12}	A_{13}	A_{14}	A_{15}	A_{16}	A_{17}	A_{18}	A_{19}
"Video 1"	73	47	78	72	91	97	76	77	54
"Video 2"	89	72	44	17	27	91	76	82	93
"Video 3"	45	19	41	89	79	97	92	17	9
"Video 4"	11	71	34	97	3	98	78	10	63
"Video 5"	29	66	71	52	13	96	61	67	32
"Video 6"	72	2	32	48	54	44	40	8	100
"Video 7"	12	2	18	53	4	90	52	33	50
"Video 8"	48	4	70	22	32	48	53	61	21

Se sobreentiende que cada atributo tiene definido una medida (medida base, medida derivada o indicador), la cual debe tener definida una unidad de medición, una escala de medición (ej: 0 a 100) y esta última debe pertenecer a un tipo específico de escala.

- *"Transparencias y Diapositivas"*

$Entidad/Atributo$	A_{21}	A_{22}	A_{23}
"Presentación 1"	61	34	36
"Presentación 2"	88	94	71
"Presentación 3"	76	45	1
"Presentación 4"	40	14	29
"Presentación 5"	92	20	35
"Presentación 6"	23	21	85
"Presentación 7"	50	78	38

- *"Aplicaciones multimedia educativas"*

$Entidad/Atributo$	A_{31}	A_{32}	A_{33}	A_{34}	A_{35}	A_{36}
"Multimedia 1"	90	21	82	67	55	66
"Multimedia 2"	6	0	61	25	46	56
"Multimedia 3"	46	34	0	36	40	53
"Multimedia 4"	83	82	67	24	79	88
"Multimedia 5"	71	5	49	45	94	41
"Multimedia 6"	91	97	2	7	98	57
"Multimedia 7"	57	53	16	90	51	53
"Multimedia 8"	73	86	46	92	43	76
"Multimedia 9"	69	92	20	5	49	8

- *"Páginas Web Educativas"*

$Entidad/Atributo$	A_{41}	A_{42}	A_{43}	A_{44}
"Web 1"	7	39	82	80
"Web 2"	38	62	76	84
"Web 3"	3	49	16	68
"Web 4"	49	17	91	70
"Web 5"	53	79	58	1
"Web 6"	24	11	81	18

En este punto es bueno mencionar que a cada categoría de entidad por separado se le puede aplicar un método de clasificación (Ver Sección 5.1.2), sin embargo, cuando las categorías entidad tienen un número diferente de atributos, no es posible aplicar estos métodos a todas las categorías simultáneamente.

6.4. Mecanismo de valoración de los indicadores

Lo ideal es que la institución disponga de un servicio web que le permita, preferiblemente, a un equipo de personas competentes medir cada uno de los objetos que van a ingresar al repositorio. Esto quiere decir que los autores no ingresan directamente al repositorio sus objetos, sino que los envían primero al dicho equipo. Otra estrategia, pero menos confiable, es que cada autor de un objeto de aprendizaje debe autoevaluar su aporte antes de adicionarlo al repositorio.

Por otro lado, la institución también debe disponer de un equipo de mejoramiento que ejecute el algoritmo de clasificación de objetos cada vez que lo requiera para identificar los mejores candidatos a ser mejorados (los de menor porcentaje de aceptación), una vez hecha esta labor, los retira del repositorio, los mejora y los ingresa nuevamente. Estas actividades de retirar e ingresar objetos al repositorio afectan, como ya se ha dicho, muy probablemente las aglomeraciones encontradas antes de realizar estas actividades.

6.5. Ejemplo del método de clasificación

6.5.1. Condiciones iniciales

- $R = \emptyset$, el repositorio debe estar totalmente vacio.
- Categorías de entidad.

$$
\begin{array}{lll}
T = \{ & (\text{“}\textit{Video didácticos}\text{”}, 0, & \\
 & \quad \{ \ (11, A_{11}, 0),\ (12, A_{12}, 0), (13, A_{13}, 0),\ (14, A_{14}, 0), (15, A_{15}, 0), & \\
 & \qquad (16, A_{16}, 0),\ (17, A_{17}, 0),\ (18, A_{18}, 0), (19, A_{19}, 0) \quad \} & \\
 &), & \\
 & (\text{“}\textit{Transparencias y Diapositivas}\text{”}, 0, & \\
 & \quad \{ \quad (21, A_{21}, 0),\ (22, A_{22}, 0), (23, A_{23}, 0) \quad \} & \\
 &), & \\
 & (\text{“}\textit{Aplicaciones multimedia educativas}\text{”}, 0, & \\
 & \quad \{ \quad (31, A_{31}, 0),\ (32, A_{32}, 0), (33, A_{33}, 0),\ (34, A_{34}, 0), & \\
 & \qquad (35, A_{35}, 0),\ (36, A_{35}, 0) \quad \} & \\
 &), & \\
 & (\text{“}\textit{Páginas Web Educativas}\text{”}, 0, & \\
 & \quad \{ \quad (41, A_{41}, 0),\ (42, A_{42}, 0), (43, A_{43}, 0),\ (44, A_{44}, 0) \quad \} & \\
 &) & \\
\}
\end{array}
$$

6.5.2. Objetos de aprendizaje

Un Objeto de aprendizaje, llamado $Entidad$ en el contexto abstracto, se define (según la Definición 6.4) como la terna $\langle C, S, P\rangle$ en donde C es el contenido del objeto de aprendizaje, S es el conjunto de los valores que toman los atributos definidos en el metadato asociado al objeto y P es el conjunto de todos los pares ordenados (i, v_i) en donde i es el código que identifica al indicador y v_i es el valor asociado a él. Para no repetir innecesariamente los resultados de la evaluación de todos los objetos de aprendizaje que aparecen desde la página 74, se va a ilustrar sólo para el objeto de aprendizaje llamado *"Presentación 1"*.

$$\textit{"Presentación 1"} = \langle C, S, \{\ (21,\ 61),\ (22,\ 34),\ (23,\ 36)\ \}\rangle$$

6.5.3. Repositorio de objetos de aprendizaje

Una vez creados los objetos de aprendizaje (Entidades), estos se insertan en el repositorio R.

$R = \{$ *"Video 1", "Video 2", "Video 3", "Video 4",*
"Video 5", "Video 6", "Video 7", "Video 8",
"Presentación 1", "Presentación 2", "Presentación 3", "Presentación 4",
"Presentación 5", "Presentación 6", "Presentación 7",
"Multimedia 1", "Multimedia 2", "Multimedia 3", "Multimedia 4",
"Multimedia 5", "Multimedia 6", "Multimedia 7", "Multimedia 8", "Multimedia 9",
"Web 1", "Web 2", "Web 3", "Web 4", "Web 5", "Web 6"
$\}$

Una vez insertados todos los objetos de aprendizaje tal y como se indica en la Sección 6.2.2, las categorías de entidad quedan así:

$T = \{$ (*"Video didácticos"*, 8,
$\{\ (11, A_{11}, 379),\ (12, A_{12}, 283), (13, A_{13}, 388),\ (14, A_{14}, 450), (15, A_{15}, 303),$
$(16, A_{16}, 661),\ (17, A_{17}, 528),\ (18, A_{18}, 355), (19, A_{19}, 422)\ \}$
),
(*"Transparencias y Diapositivas"* , 7,
$\{\ (21, A_{21}, 430),\ (22, A_{22}, 306), (23, A_{23}, 295)\ \}$
),
(*"Aplicaciones multimedia educativas"*, 9,
$\{\ (31, A_{31}, 586),\ (32, A_{32}, 510), (33, A_{33}, 343),\ (34, A_{34}, 391),$
$(35, A_{35}, 555),\ (36, A_{35}, 498)\ \}$
),
(*"Páginas Web Educativas"*, 6,
$\{\ (41, A_{41}, 174),\ (42, A_{42}, 257), (43, A_{43}, 404),\ (44, A_{44}, 321)\ \}$
)
$\}$

6.5.4. Aglomeraciones de los objetos de aprendizaje

Una vez se tiene lleno el repositorio, se procede a construir la secuencia ordenada L, tal que $\forall x \in R,\ x \in L$ y que $\forall l_i, l_j \in L$ en donde $i < j$ se debe cumplir que $\mu(l_1) \leq \mu(l_2)$.

Para este caso particular se trabajó con una cifra decimal, lo que produce máximo 11 aglomeraciones.

$$\{\mu(\circ) = 0{,}0,\ \mu(\circ) = 0{,}1,\ \mu(\circ) = 0{,}2,\ \mu(\circ) = 0{,}3,\ \mu(\circ) = 0{,}4,\ \mu(\circ) = 0{,}5,$$

$$\mu(\circ) = 0{,}6,\ \mu(\circ) = 0{,}7,\ \mu(\circ) = 0{,}8,\ \mu(\circ) = 0{,}9,\ \mu(\circ) = 1{,}0\}$$

Donde $\circ$ representa cualquier miembro de la aglomeración.

Entre más cifras decimales considere, mayor será el número de aglomeraciones y más fina será la clasificación de los objetos. Para este caso teórico y para muchos casos prácticos, una cifra decimal es más que suficiente.

$Aglomeración_1$ $\mu(*) = 0{,}0$	*"Presentación 1"*
	"Presentación 4"
	"Multimedia 3"

$Aglomeración_2$ $\mu(\circ) = 0{,}1$	*"Video 7"*
	"Multimedia 7"

$Aglomeración_3$ $\mu(\circ) = 0{,}2$	*"Multimedia 2"*
	"Multimedia 9"

$Aglomeración_4$ $\mu(\circ) = 0{,}3$	*"Web 6"*
	"Video 6"
	"Video 8"
	"Presentación 5"
	"Presentación 6"
	"Presentación 7"

$Aglomeración_5$ $\mu(\circ) = 0{,}4$	*"Video 3"*
	"Video 5"
	"Multimedia 1"
	"Multimedia 5"
	"Multimedia 6"

$Aglomeración_6$ $\mu(\cdot) = 0{,}5$	*"Web 1"*
	"Web 3"
	"Web 5"

$Aglomeración_7$ $\mu(\circ) = 0{,}6$	*"Video 4"*
	"Multimedia 4"
	"Multimedia 8"

$Aglomeración_8$ $\mu(\circ) = 0{,}7$	*"Video 2"*
	"Presentación 3"

$Aglomeración_9$ $\mu(\circ) = 0{,}8$	*"Web 4"*

$Aglomeración_{10}$ $\mu(\circ) = 1{,}0$	*"Video 1"*
	"Presentación 2"
	"Web 2"

$\circ$	Representa a cualquier miembro de la aglomeración

La forma de interpretar esta clasificación, se encuentra en el valor de $\mu(\circ)$, pues este valor indica el porcentaje de aceptación. En tal caso, { *"Presentación 1", "Presentación 4", "Multimedia 3"* } son los peores objetos de aprendizaje del repositorio, pues tiene un porcentaje de aceptación de cero. Esto quiere decir que ninguno de sus indicadores se encuentra por encima del promedio (inclusive) de este indicador en la categoría de entidad a la que corresponde el objeto de aprendizaje. Observe que en esta aglomeración hay objetos de diferentes clases.

Como se puede deducir, los objetos de aprendizaje { *"Video 1", "Multimedia 2", "Web 2"* }, son los mejores (hasta ahora) en el repositorio, no quiere decir que no tengan problemas, lo que quiere decir es que sus indicadores sí se encuentran por encima del promedio (o son iguales) de este indicador en la categoría de entidad a la que corresponde el objeto de aprendizaje.

Ahora suponga que el objeto de aprendizaje *"Multimedia 3"* es retirado del repositorio, entonces la pregunta es: ¿Cómo afecta esto las aglomeraciones?, y la respuesta es, de ninguna forma (se deja al lector como ejercicio). Pero una vez el objeto de aprendizaje *"Multimedia 3*"* es mejorado, él es medido e ingresado al repositorio como:

"Multimedia 3"*$= \langle C_{33}, S_{33}, \{ (31, 97), (32, 55), (33, 81), (34, 63), (35, 76), (36, 69)\}\rangle$

$Aglomeración_1$ $\mu(\circ) = 0{,}0$	*"Presentación 1"*
	"Presentación 4"

$Aglomeración_2$ $\mu(\circ) = 0{,}1$	*"Video 7"*
	"Multimedia 2"
	"Multimedia 7"
	"Multimedia 9"

$Aglomeración_3$ $\mu(\circ) = 0{,}3$	*"Web 6"*
	"Video 6"
	"Video 8"
	"Presentación 5"
	"Presentación 6"
	"Presentación 7"
	"Multimedia 5"
	"Multimedia 6"

$Aglomeración_4$ $\mu(\circ) = 0{,}4$	*"Video 3"*
	"Video 5"
	"Multimedia 1"
	"Multimedia 8"

$Aglomeración_5$ $\mu(\circ) = 0{,}5$	*"Web 1"*
	"Web 3"
	"Web 5"

$Aglomeración_6$ $\mu(\circ) = 0{,}6$	*"Video 4"*
	"Multimedia 3*"
	"Multimedia 4"

$Aglomeración_7$ $\mu(\circ) = 0{,}7$	*"Video 2"*
	"Presentación 3"

$Aglomeración_8$ $\mu(\circ) = 0{,}8$	*"Web 4"*

$Aglomeración_9$ $\mu(\circ) = 1{,}0$	*"Video 1"*
	"Presentación 2"
	"Web 2"

$\circ$	Representa a cualquier miembro de la aglomeración

Como se puede observar, este pequeño cambio en la inserción de un nuevo objeto de aprendizaje *"Multimedia 3*"*, provocó cambio en las aglomeraciones:

- Desaparece la aglomeración asociada a $\mu(\circ) = 0{,}2$.
- Los objetos de aprendizaje *"Multimedia 2", "Multimedia 9", "Multimedia 5", "Multimedia 6" "Multimedia 8"* y *"Multimedia 3*"* cambian de aglomeración.

Entidad	$\mu(\circ)$ **anterior**	$\mu(\circ)$ **actual**
"Multimedia 2"	0,2	0,1
"Multimedia 9"	0,2	0,1
"Multimedia 5"	0,4	0,3
"Multimedia 6"	0,4	0,3
"Multimedia 8"	0,6	0,4
"Multimedia 3"*	0,0	0,6

Como se pudo ilustrar con el ejemplo, cambios en el repositorio pueden afectar las aglomeraciones de los objetos de aprendizaje. Sin embargo, tambien se puede observar que estos cambios, si los hay, se producen únicamente en la categoría o categorías que fueron afectadas con los cambios en el repositorio. Estos cambios pueden ser significativos como por ejemplo para el objeto de aprendizaje llamado *"Multimedia 3"*.

6.6. Comentario final

En el artículo "*An Evaluation of Open Source E-Learning Platforms Stressing Adaptation Issues*"[14], se presenta una evaluación de las plataformas de *e-learning* de código abierto, este documento es importante en esta tesis pues podría servir como ejemplo para utilizar el mecanismo formal de clasificación automática en un contexto diferente al de los objetos de aprendizaje. Se observa el documento se puede ver el resultado de la evaluación de nueve plataformas, pero lo que no se ve es la forma en que estas plataformas se clasifican según los valores obtenidos en la evaluación.

CONCLUSIONES Y TRABAJO FUTURO

- Se desarrolló un mecanismo formal de clasificación automática de un repositorio de objetos de aprendizaje según el valor de un grupo de indicadores (Capítulo 6).

  - Se construyó un espacio de medida (la terna (Conjunto de objetos X, la sigma álgebra Σ, una función de medida μ)) que sirve para clasificar automáticamente un conjunto de objetos de aprendizaje según el valor numérico de unos indicadores (Sección 6.1).
  - Se logró demostrar matemáticamente que la función de medida μdefinida es una función válida (Teorema 6.1 en la Sección 6.1).
  - Se logró demostrar matemáticamente que la función de disimilaridad definida satisface todas la propiedades (Teorema 6.2 en la Sección 6.1).
  - Se definió un método de clasificación que utilice la medida definida (Sección 6.2).
  - Se analizaron los requerimientos que deben cumplir los algoritmos de aglomeramiento (Sección 5.1) con respecto a la nueva propuesta y se pudieron observar sus bondades.
  - Se determinaron los indicadores para clasificar los objetos de aprendizaje en un contexto particular, en este caso un constexto abstracto (Sección 6.3).
  - Se estableció un mecanismo de valoración de los indicadores definidos (Sección 6.4).
  - Se desarrolló un ejemplo del método de clasificación para el contexto particular definido (Sección 6.5).

- Para este mecanismo de clasificación en particular, se consideran sólo los indicadores que puedan ser valorados numéricamente (Ver primer objetivo específico de la página 30). Sin embargo, esto no debe ser visto como una debilidad, por el contrario, debe ser visto como una oportunidad para futuras investigaciones. La idea, sería tomar la información de la Sección 5.1.1 en la página 39 y crear una nueva función de medida, que por supuesto hay que demostrar matemáticamente, tal y como se hizo con la función propuesta.

- Como se pudo mostrar con esta tesis, es posible construir mecanismos automáticos de clasificación de un conjunto de objetos de aprendizaje. Aún si estos objetos son de categorías diferentes. Es más, este mecanismo de clasificación puede ser utilizado para diferentes tipos de elementos no necesariamente objetos de aprendizaje.

- Sería deseable utilizar siempre un proceso de medición como el estudiado en la Sección 5.3, puesto que esto ayuda a dar mayor significancia a la información obtenida por cada objeto y por las posibles aglomeraciones obtenidas como producto de la clasificación.

BIBLIOGRAFÍA

[1] Anderberg, Michael R., "Cluster analysis for applications", Academic Press, 1973

[2] Berkhin Pavel, "Survey of clustering data mining techniques", 2002. [Sitio en Internet]. Disponible en http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.18.3739, Acceso en marzo de 2009.

[3] Boneu Josep M., "Plataformas abiertas de e-learning para el soporte de contenidos educativos abiertos". [Sitio en Internet]. Disponible en http://www.uoc.edu/rusc/4/1/dt/esp/boneu.pdf, ,Universitat Oberta de Catalunya (UOC), Acceso en abril de 2009.

[4] C. M Benjamin, Ke Fung , Ester Wang Martin, "Large hierarchical document clustering using frequent itemsets", 2003. [Sitio en Internet]. Disponible en http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.13.9326, Acceso en marzo de 2009.

[5] Cao Zhe, Qin Tao, Liu Tie-Yan, Tsai Ming-Feng, Li Hang, "Learning to Rank: From Pairwise Approach to Listwise Approach", 2007. [Sitio en Internet]. Disponible en http://research.microsoft.com/en-us/people/tyliu/139.pdf, Acceso en octubre de 2009.

[6] Chakrabarti Soumen, "Data mining for hypertext: A tutorial survey", 2000. [Sitio en Internet]. Disponible en http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.34.4297, Acceso en marzo de 2009.

[7] Chechik Gal, Sharma Varun, Shalit Uri, Bengio Samy, "Large Scale Online Learning of Image Similarity Through Ranking", 2009. [Sitio en Internet]. Disponible en http://snowbird.djvuzone.org/abstracts/119.pdf, Acceso en octubre de 2009.

[8] Cormen Thomas H., Leiserson Charles E., Rivest Ronald L., Stein Clifford, "Introduction to Algorithms", Second Edition, The MIT Press, McGraw-Hill, 2001.

[9] Cui Xiaohui, Potok Thomas E., "Document Clustering Analysis Based on Hybrid PSO+K-means Algorithm", 2005. [Sitio en Internet]. Disponible en http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.127.8448, Acceso en marzo de 2009.

[10] Driver Bruce K., "Analysis Tools with Applications", Springer, 2003. Pag. 25-26, 131-135.

[11] Deshpande Mukund, Karypis George, "Using Conjunction of Attribute Values for Classification", 2002. [Sitio en Internet]. Disponible en http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.16.6905, Acceso en marzo de 2009.

[12] desJardins Marie, Eaton Eric, Wagstaff Kiri L., "Learning User Preferences for Sets of Objects", 2006. [Sitio en Internet]. Disponible en http://maple.cs.umbc.edu/~ericeaton/papers/desJardins2006Learning.pdf, Acceso en octubre de 2009.

[13] Ferreira Mateus, García Félix, Ruiz Francisco, Bertoa Manuel F., Calero Coral, Vallecillo Antonio, Piattini Mario, Mora Beatriz, "Medición del Software Ontología y Metamodelo (Informe Técnico UCLM-TSI-001)", Departamento de Tecnologías y Sistemas de Información, Universidad de Castilla-La Mancha y Universidad de Málaga, España, 2006. [Sitio en Internet]. Disponible en http://www.esi.uclm.es:8080/tsi/informes/UCLM-TSI-001.pdf, Acceso en abril de 2009.

[14] Graf Sabine, List Beate, "An Evaluation of Open Source E-Learning Platforms Stressing Adaptation Issues", 2005. [Sitio en Internet]. Disponible en http://wit.tuwien.ac.at/people/list/publications/icalt2005.pdf, Acceso en octubre de 2009.

[15] Han Jiawei, Kamber Micheline, Data Mining, "Concepts and Techniques", Second Edition, Morgan Kaufmann Publishres, 2006, Chapter 7 "Cluster Analysis", Pag. 383-464

[16] Hammouda Khaled M. , Member Student , Kamel Mohamed S. , "Efficient phrase-based document indexing for Web document clustering", 2004. [Sitio en Internet]. Disponible en http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.85.6184, Acceso en marzo de 2009.

[17] Hofmann Thomas, Puzicha Jan, "Data mining for hypertext: A tutorial survey", 2000. [Sitio en Internet]. Disponible en http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.34.97, Acceso en marzo de 2009.

[18] IEEE LOM Institute for Electrical and Electronic Engineers Standards Department, "Learning Object Metadata (IEEE 1481.12.1-2002)", 2002. [Sitio en Internet]. Disponible en http://ieeexplore.ieee.org/xpl/standards.jsp?findtitle=learning+object+metadata&letter=learning+object+metadata&imageField.x=0&imageField.y=0&opentree=on, o también se puede acceder en español en http://www-gist.det.uvigo.es/~lanido/LOMes/LOMv1_0_Spanish.pdf Acceso en mayo 2008

[19] Kaur Kamaldeep, Kaur Arvinder, Malhotra Ruchika, "Alternative Methods to Rank the Impact of Object Oriented Metrics in Fault Prediction Modeling using Neural Networks", 2006, [Sitio en Internet]. Disponible en http://www.waset.org/journals/waset/v19/v19-27.pdf, Acceso en octubre 2009.

[20] Khan Inayatullah, Saffari Amir, Bischof Horst, "TVGraz: Multi-Modal Learning of Object Categories by Combining Textual and Visual Features", 2009, [Sitio en Internet]. Disponible en http://oagm2009.icg.tugraz.at/papers/p36.pdf, Acceso en octubre 2009.

[21] Li Hang, "Learning to Rank:Problem, Challenge, and Opportunity", Microsoft Research Asia, 2008, [Sitio en Internet]. Disponible en http://research.microsoft.com/users/tyliu/, Acceso en octubre 2009.

[22] Liu Tie-Yan, "Learning to Rank: from PairwiseApproach to ListwiseApproach", Microsoft Research Asia, 2008, [Sitio en Internet]. Disponible en http://research.microsoft.com, Acceso en octubre 2009.

[23] Marquès Pere, "Sitio Web sobre Tecnología Educativa", [Sitio en Internet]. Disponible en http://www.pangea.org/peremarques/, Acceso en julio 2009.

[24] Mimaroglu Selim, Simovici Dan A., "Bit Sequences and Biclustering of Text Documents", [Sitio en Internet] http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.132.4386, Acceso en marzo de 2009.

[25] Nesbit John C., Li Jerry, "Web-Based Tools for Learning Object Evaluation", 2004, [Sitio en Internet] http://www.sfu.ca/~jzli/publications/Nesbit_Li_2004.pdf, Acceso en octubre de 2009.

[26] Nesbit John C., Li Jerry, Leacock Tracey L., "Web-Based Tools for Collaborative Evaluation of Learning Resources", 2004, [Sitio en Internet]. Disponible en http://www.iiisci.org/journal/CV$/sci/pdfs/P843017.pdf, Acceso en octubre 2009.

[27] Nie Zaiqing, Zhang Yuanzhi, Wen Ji Rong, Ma WeiYing, "Object Level Ranking: Bringing Order to Web Objects", 2005, [Sitio en Internet]. Disponible en http://www2005.org/cdrom/docs/p567.pdf, Acceso en octubre 2009.

[28] "Proceedings of the NIPS 2005 Workshop on Learning to Rank", Whistler, BC, Canada, 2005, [Sitio en Internet]. Disponible en http://web.mit.edu/shivani/www/Ranking-NIPS-05/Proceedings/proceedings-nips05workshop-ranking.pdf, Acceso en octubre 2009.

[29] Ochoa Xavier, "The research about Metrics for Metadata", Hypermedia and Databases Group (HMDB) of the Katholieke Universiteit Leuven (KULeuven), 2006. [Sitio en Internet]. Disponible en http://ariadne.cti.espol.edu.ec/M4M/index.html, Acceso en marzo de 2009.

[30] Ochoa Xavier, Duval Erik, "Metrics for Learning Object Metadata", 2006. [Sitio en Internet]. Disponible en http://ariadne.cti.espol.edu.ec/xavier/papers/Ochoa-DC-ECTEL2006.pdf, Acceso en octubre de 2009.

[31] Ochoa Xavier, Duval Erik, "Use of contextualized attention metadata for ranking and recommending learning objects", 2006. [Sitio en Internet]. Disponible en http://portal.acm.org/ft_gateway.cfm?id=1183608&type=pdf&coll=GUIDE&dl=GUIDE&CFID=52920346&CFTOKEN=27345131, Acceso en octubre de 2009.

[32] Ochoa Xavier, Duval Erik, Quantitative Analysis of Learning Object Repositories", 2008. [Sitio en Internet]. Disponible en http://bobbebaggio.com/Presentations/PADLA111208/ROL-Repositories.pdf, Acceso en octubre de 2009.

[33] Özgür Arzucan, Güngör Tunga, "Supervised and Unsupervised Machine Learning Techniques for Text Document Categorization", 2004. [Sitio en Internet]. Disponible en http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.123.1804, Acceso en marzo de 2009.

[34] Parakhin Mikhail, Haluptzok Patrick, "Finding The Most Probable Ranking of Objects with Probabilistic Pairwise Preferences",2009. [Sitio en Internet]. Disponible en http://www.cvc.uab.es/icdar2009/papers/3725a616.pdf, Acceso en octubre de 2009.

[35] Peters Gabriele, Kruger Norbert, Von Der MalsburgChristoph, "Learning object representations by clustering banana wavelet responses",1997. [Sitio en Internet]. Disponible en http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.50.4104, Acceso en marzo de 2009.

[36] Qin Tao, Liu Tie-Yan, Zhang Xu-Dong, Wang De-Sheng, Xiong Wen-Ying, Li Hang,"Learning to Rank Relational Objects and Its Application to Web Search", 2008, [Sitio en Internet] http://www2008.org/papers/pdf/p407-qinA.pdf, Acceso en octubre de 2009.

[37] Repp Stephan, Linckels Serge, Meinel Christoph, "Towards to an automatic semantic annotation for multimedia learning objects",2007. [Sitio en Internet]. Disponible en http://portal.acm.org/ft_gateway.cfm?id=1290149&type=pdf&coll=GUIDE&dl=GUIDE&CFID=52920346&CFTOKEN=27345131, Acceso en marzo de 2009.

[38] Rosales Rómer, Fung Glenn, "Learning sparse metrics via linear programming", 2006. [Sitio en Internet]. Disponible en http://portal.acm.org/ft_gateway.cfm?id=1150444&type=pdf&coll=ACM&dl=ACM&CFID=52920346&CFTOKEN=27345131, Acceso en octubre de 2009.

[39] Seidl Thomas, Muller Emmanuel, Assent Ira, Steinhausen Uwe, "Outlier Detection and Ranking Based on Subspace Clustering", 2009,[Sitio en Internet] http://drops.dagstuhl.de/opus/volltexte/2009/1934/pdf/08421.SeidlThomas.Paper.1934.pdf, Acceso en octubre de 2009.

[40] Stephenson Jhon, Sangrà Albert, "Fundamentos del diseño técnico-pedagógico", P06/M1103/01178, Universitat Oberta de Catalunya (UOC).

[41] Strehl Alexander, Ghosh Joydeep, "Relationship-based clustering and visualization for high-dimensional data mining", 2003,[Sitio en Internet] http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.19.7967, Acceso en marzo de 2009.

[42] Shaban Khaled, Basir Otman, Kamel Mohamed, "Learning Objects Clustering based on Semantic Understanding of Text", 2006, [Sitio en Internet] http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.132.428, Acceso en marzo de 2009.

[43] Vidal Christian L., Segura Alejandra A., Prieto Manuel E.,"Calidad en objetos de aprendizaje", 2008, [Sitio en Internet] http://www.web.upsa.es/spdece08/contribuciones/139_CalidadEnObjetosDeAprendizaje TypeInstSpringerFinalVidalSeguraPrietov99.pdf, Acceso en octubre de 2009.

[44] Wiley David, "Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy". The Instructional Use of Learning Objects: Online Versión. 2000. [Sitio en Internet] http://reusability.org/read/chapters/wiley.doc, Acceso en febrero de 2009.

[45] Woznica Adam, Kalousis Alexandros, Hilario Melanie, "Learning to combine distances for complex representations". 2007. [Sitio en Internet] http://portal.acm.org/ft_gateway.cfm?id=1273626&type=pdf&coll=

GUIDE&dl=GUIDE&CFID=52920346&CFTOKEN=27345131, Acceso en octubre de 2009.

[46] Yeh Jen-Yuan, Lin Jung-Yi, Ke Hao-Ren, Yang Wei-Pang, "Learning to Rank for Information Retrieval Using Genetic Programming", 2007, [Sitio en Internet] http://jenyuan.yeh.googlepages.com/jyyeh-LR4IR07.pdf, Acceso en octubre de 2009.

[47] Zhao Ying, Karypis George, "Soft Clustering Criterion Functions for Partitional Document Clustering", 2002, [Sitio en Internet] http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.5.8340, Acceso en marzo de 2009.

[48] Zhao Ying, Karypis George, "Criterion functions for document clustering: Experiments and analysis", 2001, [Sitio en Internet] http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.16.6872, Acceso en marzo de 2009.