PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
ESCUELA DE INGENIERÍA – FACULTAD DE LETRAS – BIBLIOTECAS UC

EXPLORACIÓN DE LA RELACIÓN ENTRE RENDIMIENTO ACADÉMICO DE ALUMNOS DE PREGRADO, CONSULTAS DE LOS SERVICIOS DE BIBLIOTECA Y MULTIDISCIPLINARIEDAD, APLICANDO TÉCNICAS DE MINERÍA DE DATOS.

MARIBEL ALVARADO ACUÑA

Proyecto de Tesis para optar al Grado de
Magíster en Procesamiento y Gestión de la Información

Profesor Supervisor:
DENIS PARRA SANTANDER

Santiago de Chile, Mayo, 2016.

PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
ESCUELA DE INGENIERÍA – FACULTAD DE LETRAS – BIBLIOTECAS UC

EXPLORACIÓN DE LA RELACIÓN ENTRE RENDIMIENTO ACADÉMICO DE ALUMNOS DE PREGRADO, CONSULTAS DE LOS SERVICIOS DE BIBLIOTECA Y MULTIDISCIPLINARIEDAD, APLICANDO TÉCNICAS DE MINERÍA DE DATOS.

MARIBEL ALVARADO ACUÑA

Proyecto de Tesis presentado a la Comisión integrada por:

DENIS PARRA SANTANDER

JORGE GANA LEAY

ELVIRA SAURINA DE SOLANES

MAXIMILIANO MONTENEGRO MAGGIO

Santiago de Chile, Mayo, 2016.

Dedicatoria

A Dios, mi familia y amigos.

Quienes me acompañaron durante todo este camino.

Especialmente a Luna, Natalia y Ricardo.

Por el tiempo restado.

Agradecimientos

Mis agradecimientos a Bibliotecas UC, a Evelyn Didier y María Luisa Arenas, por darme la oportunidad de realizar mis estudios de Magister.

Al equipo del Proyecto Fondecyt 1131109: “Learning analytics: understanding relationships between eLearning and library use, student perceptions of these tools and learning outcomes”, Magdalena Jara, Elvira Saurina, Rosa Alarcón, Carlos González, Maximiliano Montenegro, Camila León, Augusto Sandoval y Paula Clasing, por invitarme a colaborar en una fructífera experiencia de trabajo interdisciplinar.

A la Unidad de Tecnología de la Información de Bibliotecas UC, mis compañeros de trabajo, especialmente a Alejandro Burgos por su constante ayuda y por entregarme la confianza necesaria para realizar esta investigación.

Al programa de Magister en Procesamiento y Gestión de la Información, de la Pontificia Universidad Católica de Chile, por entregarme las pautas y herramientas para replantear mi profesión en una forma distinta.

A mi profesor guía, Denis Parra, por sus enseñanzas y orientaciones para desarrollar mi tesis.

A mi hermano, Luis Alvarado, quién compartió conmigo horas de su apretada agenda para darme consejos y sugerir metodologías.

Finalmente a mis amadas hijas y esposo, por su paciencia para aceptar mi tiempo reducido y apoyarme durante estos dos años de estudio.

Tabla de contenido

Dedicatoria 2
Agradecimientos 3
Tabla de contenido 4
Índice de tablas 6
Índice de figuras 7
Resumen 9
Abstract 11
Introducción 13
1. Formulación del problema 18
1.1 Preguntas para la investigación 20
2. Objetivos 21
Objetivo de la investigación 21
Objetivo general 21
Objetivos específicos 21
3. Motivación/Justificación: 22
4. Marco teórico y conceptual 24
4.1 Marco teórico 24
4.2 Marco conceptual 35
5. Hipótesis 39
6. Aspectos metodológicos 40
6.1 Metodología 40
6.2 Recolección de Datos 42
6.3 Descripción de los conjuntos de datos 43
7. Resultados 48
7.1 Consultas de la colección presencial 48
7.1.1 Análisis estadísticos 48
7.1.2 Conjuntos frecuentes de áreas consultadas 52
7.1.3 Entropía de las consultas por área 54
7.1.4 Reglas de asociación por áreas 55
7.1.5 Consultas por área del material y área del usuario 58

7.1.6 Uso de la biblioteca y notas 64
7.1.7 Notas y cantidad de áreas consultadas 68
7.1.8 Clustering de transacciones y notas 79
7.2 Consultas a la colección electrónica 85
7.2.1 Análisis estadístico 85
7.2.2 Consultas por área y carrera del usuario 87
7.2.3 Consultas a la colección electrónica y uso de la biblioteca 90
7.2.4 Consultas a la colección electrónica y notas 96
7.2.5 Análisis de dominios consultados 100
7.2.6 Conjuntos frecuentes de dominios consultados 103
7.2.7 Reglas de asociación por dominio 104
7.2.8 Clustering de consultas y notas 107
8. Conclusiones 120
8.1 Colección presencial 120
8.2 Consultas a la colección electrónica 124
8.3 Contrastación de resultados con hipótesis 128
8.4 Sugerencias 134
Bibliografía consultada 136
Anexos 140
Anexo 1: Consultas SQL utilizadas en tablas de Aleph 140
Anexo 2: Tipos de usuarios definidos en la Biblioteca 141
Anexo 3: Áreas temáticas y Áreas OCDE 142
Anexo 4: Reglas de asociación 2012-2014 145

Índice de tablas

Tabla 6-1: Cantidad de Consultas de la colección presencial 2010-2014 43
Tabla 6-2: Matriz de Consultas de la colección presencial 2010-2014 44
Tabla 6-3: Matriz de rendimiento académico y transacciones 2012-2014 45
Tabla 6-4: Matriz de áreas del material, carrera y zscore del alumno 2012-2014 46
Tabla 6-5: Matriz de Zscore y EzProxy, sesiones y consultas, 2012-2014 47
Tabla 7-1: Itemsets frecuentes 2010-2014 53
Tabla 7-2: Reglas de asociación por áreas 56
Tabla 7-3: Promedio de consultas por área del material y área del usuario 2012-2014 ... 58
Tabla 7-4: Cantidad de áreas consultadas por área del usuario 59
Tabla 7-5: Test binomial de consultas en áreas distintas a la de la Unidad Académica ... 62
Tabla 7-6: Proporción de multidisciplinariedad de las consultas 63
Tabla 7-7: Prueba de proporciones entre promedios de notas y promedios de consultas 67
Tabla 7-8: Tabla de proporciones entre notas y cantidad de áreas consultadas 73
Tabla 7-9: Tabla de proporciones entre notas y cantidad de áreas consultadas por área 74
Tabla 7-10: Análisis estadístico de notas y cantidad de áreas consultadas 76
Tabla 7-11: Regresión lineal de notas y áreas consultadas por los alumnos 77
Tabla 7-12: Prueba t de notas y áreas consultadas por los alumnos 77
Tabla 7-13: Prueba de proporciones para cantidad de áreas consultadas y notas 78
Tabla 7-14: Vectores de desempeño de los clusters con K-means 81
Tabla 7-15: Vector de desempeño de los clusters con DBSCAN 84
Tabla 7-16: Cantidad de usuarios de la colección electrónica 85
Tabla 7-17: Cantidad de sesiones de consulta de la colección electrónica 85
Tabla 7-18: Duración de las sesiones de consultas a la colección electrónica 86
Tabla 7-19: Promedio de minutos por sesión y por área de los usuarios 88
Tabla 7-20: % de crecimiento de las consultas y su duración 2012-2014 89
Tabla 7-21: Correlación entre variables Zscore, sesiones, duración y uso de biblioteca .. 92
Tabla 7-22: Análisis estadístico de Sesiones y uso de la biblioteca 2012 a 2014 93
Tabla 7-23: Correlación para cantidad de sesiones y uso de biblioteca 94
Tabla 7-24: Regresión lineal para sesiones y notas 97
Tabla 7-25: Prueba de proporciones entre Cantidad de sesiones y Zscore de notas 98
Tabla 7-26: Prueba de proporciones entre duración de las sesiones y notas 99
Tabla 7-27: Conjunto frecuentes de dominios consultados 103
Tabla 7-28: Reglas de asociación por dominios consultados 2012 105
Tabla 7-29: Reglas de asociación por dominios consultados 2013 105
Tabla 7-30: Reglas de asociación por dominios consultados 2014 106
Tabla 7-31: Vector de desempeño de los clusters con K-means 114
Tabla 7-32: Vector de desempeño para clusters con Dbscan 119
Tabla 8-1: Análisis estadístico de la multidisciplinariedad de las consultas 129
Tabla 8-2: Uso de biblioteca y notas 2012-2014 130
Tabla 8-3: Uso de colección electrónica y notas 2012-2014 131
Tabla 8-4: Duración de sesiones y notas 2012-2014 131
Tabla 8-5: Correlación entre Consultas presenciales y a la colección electrónica. 133

Índice de figuras

Imagen 4-1: Proceso de descubrimiento de conocimiento 27
Imagen 4-2: Proceso de minería de datos 27
Imagen 7-1: Transacciones por área 48
Imagen 7-2: Préstamos promedio por área 48
Imagen 7-3: Usuarios que piden en un área o más 49
Imagen 7-4: Cantidad de áreas consultadas durante el 2010 al 2014, en %. 50
Imagen 7-5: Consultas por área y combinación de áreas 50
Imagen 7-6: Transacciones de alumnos regulares por área 51
Imagen 7-7: Transacciones de usuarios no estudiantes por área 52
Imagen 7-8: Entropía por áreas 54
Imagen 7-9: Soporte de reglas de asociación 57
Imagen 7-10: Transacciones promedio por área del usuario y área del material 59
Imagen 7-11: Cantidad de áreas consultadas por área del usuario 60
Imagen 7-12: Cantidad de áreas consultadas por carrera 61
Imagen 7-13: Promedio de transacciones y Zscore por carrera 2012-2014 64
Imagen 7-14: Promedio de transacciones por área de usuario 65
Imagen 7-15: Dispersión de promedio de transacciones y Zscore 66
Imagen 7-16: Cantidad de áreas consultadas y zscore de notas año 2012 68
Imagen 7-17: Multidisciplinariedad de consultas por área del usuario y notas año 2012 .. 69
Imagen 7-18: Cantidad de áreas consultadas y zscore de notas año 2013 70
Imagen 7-19: Multidisciplinariedad de consultas por área del usuario y notas año 2013 .. 70
Imagen 7-20: Cantidad de áreas consultadas y zscore de notas año 2014 71
Imagen 7-21: Multidisciplinariedad de consultas por área del usuario y notas año 2014 .. 72
Imagen 7-22: Multidisciplinariedad de consultas por área del usuario y notas 2012-2014 72
Imagen 7-23: Multidisciplinariedad y notas por áreas de los usuarios, años 2012-2014 .. 73
Imagen 7-24: Correlación entre notas y cantidad de áreas consultadas años 2012-2014 75
Imagen 7-25: Notas y cantidad de áreas consultadas durante los años 2012 y 2014 75
Imagen 7-26: Clusters de transacciones y Zscore con K-means para 2012 80
Imagen 7-27: Clusters de transacciones y Zscore con K-means para 2013 80
Imagen 7-28: Clusters de transacciones y Zscore con K-means para 2014 80
Imagen 7-29: Clusters de transacciones y Zscore con DBSCAN para 2012 82
Imagen 7-30: Clusters de transacciones y Zscore con DBSCAN para 2013 83
Imagen 7-31: Clusters de transacciones y Zscore con DBSCAN para 2014 83
Imagen 7-32: Promedio de duración de consultas a la colección electrónica 86
Imagen 7-33: Promedio de sesiones en EzProxy por área del usuario 87
Imagen 7-34: Promedio de sesiones y duración de las sesiones por área del usuario 88
Imagen 7-35: Promedio de sesiones y consultas por carrera 89
Imagen 7-36: Consultas presenciales y sesiones por año de la carrera 90
Imagen 7-37: Sesiones y Zscore de notas por año de carrera 91
Imagen 7-38: Promedio de consultas por Zscore 91

Imagen 7-39: Transacciones en biblioteca y consultas EzProxy por año de carrera 92
Imagen 7-40: Cantidad de sesiones y Zscore de notas obtenidas por año 96
Imagen 7-41: Cantidad de sesiones promedio por zscore de notas. 97
Imagen 7-42: Cantidad de consultas y dominios consultados 100
Imagen 7-43: Consultas por dominios 101
Imagen 7-44: Consultas por dominio en detalle 102
Imagen 7-45: Clusters de notas, sesiones y duración con K-means para 2012 108
Imagen 7-46: Clusters de notas, sesiones y duración con K-means para 2013 110
Imagen 7-47: Clusters de notas, sesiones y duración con K-means para 2014 112
Imagen 7-48: Clusters de notas, sesiones y duración con DBSCAN para 2012 115
Imagen 7-49: Clusters de notas, sesiones y duración con DBSCAN para 2013 117
Imagen 7-50: Clusters de notas, sesiones y duración con DBSCAN para 2014 118
Imagen 8—8-1: Grupos obtenidos en clusters de transacciones presenciales y notas .. 123
Imagen 8—8-2: Grupos obtenidos en clusters de sesiones y notas 126
Imagen 8—8-3: Grupos obtenidos en los clusters de duración de sesiones y notas 127
Imagen 8—8-4: Grupos obtenidos en los clusters de sesiones y duración 127

Resumen

Se explora la relación entre el rendimiento académico de los alumnos de pregrado de la Pontificia Universidad Católica de Chile, el uso de las colecciones de Bibliotecas UC y la multidisciplinariedad de las consultas, aplicando técnicas de minería de datos. Este estudio es una de las primeras experiencias en Chile para conocer como los factores relacionados con el uso de los servicios de biblioteca influyen en el rendimiento académico de los estudiantes.

Se busca identificar y explorar los perfiles de usuarios cuyos préstamos de material de biblioteca reflejan consultas multidisciplinarias, para responder la primera hipótesis del estudio "Menos del 50% de las consultas son multidisciplinarias".

Luego se analiza la relación entre el uso de los servicios de la biblioteca, presencial y electrónica, y el rendimiento de los alumnos, con una evaluación en el tiempo, para responder a la segunda hipótesis de este estudio "Los alumnos con mayor rendimiento son los que usan los servicios de la biblioteca por sobre el promedio", y aportar a una de las hipótesis del proyecto Fondecyt no. 1131109 en cuanto a determinar "Cuáles son las métricas apropiadas para establecer relaciones entre uso de biblioteca, tecnología de investigación electrónica y resultados de aprendizajes".

Adicionalmente se compara, con una evaluación en el tiempo, las consultas en la biblioteca y a la colección electrónica, con el fin de definir la tercera hipótesis del estudio "Existe relación entre las consultas de la colección electrónica y el uso de la biblioteca presencial".

Se trabajó con datos de las transacciones en los servicios presenciales, entre los años 2010-2014, datos de consulta a la colección electrónica y de rendimiento de los alumnos, entre los años 2012-2014. Se utiliza técnicas de análisis estadístico y se aplican modelos de asociación y clustering.

Los resultados indican que, respecto al uso de la colección presencial, se observa una disminución de las transacciones en todas las áreas, con una disminución mayor en aquellas áreas que tradicionalmente presentan mayor cantidad de transacciones. Existe una tendencia a la consulta en una o dos áreas, las consultas multidisciplinarias son menos

que las monodisciplinarias. Se observa una débil correlación entre la cantidad de áreas consultadas y el rendimiento académico. No existen usuarios, de carreras ni áreas, que consulten exclusivamente de su área. Se detectó tres grupos de alumnos, el grupo minoritario presenta la mayor cantidad de transacciones y las mejores notas. En general, la correlación entre notas y uso de la biblioteca es positiva: en la medida que suben las transacciones hay mejores notas, entre los usuarios que consultan por sobre el promedio de consultas la mayoría tienen notas sobre el promedio.

En relación a las consultas a la colección electrónica se observó que existe una fuerte tendencia a ir aumentando la cantidad de sesiones en el tiempo, en los usuarios de todas las áreas. Hay una correlación positiva entre cantidad de sesiones y uso de la biblioteca, la que ha ido disminuyendo en el tiempo sin llegar a ser negativa. También se detectó tres grupos, el grupo minoritario de alumnos presenta sesiones sobre la media y las mejores notas. Se observa una tendencia a tener mayor cantidad de transacciones en aquellos con mejores notas. Lo mismo sucede con la variable duración de las sesiones. Se observa además una tendencia a tener pocas sesiones y cortas, no existen casos de alumnos con una alta cantidad de sesiones y larga duración. En general, se observa lo mismo que para las transacciones presenciales, entre los usuarios que consultan por sobre el promedio de consultas la mayoría tienen notas sobre el promedio.

Abstract

The relationship between academic performance of undergraduate students at the Pontifical Catholic University of Chile (UC), the use of UC Library collections and multidisciplinary consultations, is explored applying data mining techniques. This study is one of the first experiences in Chile to know as factors related to the use of library services influencing academic performance of students.

It seeks to identify and explore users' profiles whose library material´s loans reflect multidisciplinary consultations, to answer the first hypothesis of this study: "Less than 50% of users' consults are multidisciplinaries".

Then the relationship between the use of library services, both physicals and electronics, and student performance, with an evaluation in time, is explored, to answer the second hypothesis of this study: "The students with better performance are those who consult library services over the average"; and to collaborate with one of the hypothesis of the Fondecyt Project no. 1131109. In order to determine "Which are the appropiated metrics to establish relations between library use, electronic research technology and learning outcomes".

Additionally, transactions in the library and access to the electronic collection are compared, with an evaluation in the time, in order to define the third hypothesis of the study "Exist a relation between consultations of electronic collection and use of the physical library".

It was worked with transactions' data in physical services, between the years 2010-2014, session's data to the electronic collection and student performance between the years 2012-2014. Statistical analysis techniques were used and association and clustering models were applied.

Regarding the use of physical collection, the results indicate that a decrease of consultations was observed in all areas, with a greater decrease in those areas with the highest number of transactions. There is a tendency to consult in one or two areas, the multidisciplinaries consults are lower than the monodisciplinaries. A weak correlation between the number of areas consulted and academic performance is observed. There are not users, in careers or areas that consult exclusively one area. Three groups of students were detected, the

minority group has the highest number of transactions and the best grades. Users with low grades and a high number of transactions were not observed. In general, the correlation between notes and library use is positive, as the number of transactions rise there are better grades, between the users with transactions over the average the majority have grades over the average.

.

Regarding sessions in the electronic collection, a strong tendency to increase the number of sessions over time was observed, with users in all areas. There is a positive correlation between number of sessions and library use, as long as a service is used there is a tendency to use the other, there is a complementary. Three groups were also detected, the minority group of students presented sessions on average and the best grades. A trend to have more transactions in those with better grades was observed. The same applies to the variable length of session. There is also a tendency to have few and short sessions; there are no cases of students with a high number of long sessions. In general, the same for physicals transactions is observed, the majority of users with transactions over the average have grades over the average.

Introducción

La minería de datos es un término que se comenzó a usar en la década de los 90, pero la idea de encontrar correlaciones entre variables de una base de datos ya existía desde los años ’60. Como parte del proceso de descubrimiento de conocimiento en bases de datos (KDD - Knowledge Discovery in Databases), desde sus inicios la minería de datos ha sido incrementalmente aplicada en diversas áreas como astronomía, marketing, inversiones, detecciones de fraude, manufacturas y limpieza de datos (Fayyad, Piatetsky-Shapiro & Smith, 1996). Con el paso de los años estas aplicaciones han ido ampliándose, agregándose otras áreas como industrias del retail y telecomunicaciones, sistemas recomendadores, análisis de redes sociales, comercio electrónico, seguridad y bibliotecas, por nombrar algunas.

Las bibliotecas generan enormes cantidades de datos, desde sus sistemas integrados, bases de datos, servicios presenciales y en línea. Ya a inicios de la década del 2000 se estaba realizando los primeros estudios de análisis aplicando técnicas de minería de datos sobre datos provenientes de diversas fuentes. En el mercado ya están surgiendo las primeras plataformas que pretenden unificar estas fuentes de datos y entregar análisis integrados. Las analíticas en bibliotecas se están haciendo cada vez más necesarias en las distintas áreas, generándose incluso una nueva terminología: Bibliomining, para referirse a la aplicación de minería de datos en bibliotecas.

Por otra parte, en el área de educación también han surgido interesantes estudios que utilizan técnicas de minería de datos, surgiendo lo que se denomina minería de datos educacional, el área que interesa en esta investigación específicamente corresponde a las analíticas aprendizaje o Learning Analytics.

Es en esta área en donde se busca definir las variables que se asocian al aprendizaje, en una Universidad este tópico toma una relevancia mayor ya que el nivel de los alumnos redunda en la visión que tenga la sociedad de la calidad de la casa de estudio. Los factores que influyen en el aprendizaje de los estudiantes pueden ir desde habilidades y destrezas personales, hasta la formación anterior y herramientas de las que disponen y utilizan para optimizar su proceso formativo. En este estudio se considera la biblioteca, y todos los servicios que esta ofrece, como un recurso del que disponen los estudiantes para apoyar su proceso de aprendizaje.

El presente estudio se refiere al tema del análisis de datos, específicamente a la búsqueda de patrones de uso y asociaciones entre variables, utilizando técnicas de minería de datos. La motivación para realizarlo es conocer el rol de la Biblioteca como factor influyente en el rendimiento académico de los estudiantes, y la tendencia en la multidisciplinariedad de las consultas de nuestros usuarios.

La medición de la incidencia de la biblioteca y sus servicios en el rendimiento académico de los alumnos en una universidad, permite posicionar mejor la biblioteca dentro de la universidad en la competencia por recursos económicos con las otras unidades. Si se demuestra que la biblioteca es un factor que propicia mejor rendimiento en los alumnos estará mejor posicionada para pedir apoyo financiero con el fin de optimizar sus servicios y mejorar sus colecciones con aquellos recursos que son de interés para los usuarios.

Para fomentar un desarrollo de colecciones acorde con los intereses de los usuarios, es necesario conocer cuáles son sus patrones de uso, sus áreas de interés, las tendencias en el tiempo y establecer perfiles de usuarios. En todas estas tareas los análisis realizados con técnicas de minería de datos permitirán tener una toma de decisiones más científica, lo que permitirá apoyar la gestión de la biblioteca.

El análisis de datos generados en Bibliotecas de la Pontificia Universidad Católica de Chile (Bibliotecas UC), aplicando técnicas de minería de datos, se consideró por las siguientes razones:

- Interés personal de conocer esta disciplina, como parte de la unidad de Tecnología para Bibliotecas se requiere conocer diversas metodologías de análisis para apoyar el área de gestión de la Biblioteca, generando nuevos análisis que sean útiles para la gestión de Bibliotecas UC.
- Interés académico de establecer nuevas experiencias en bibliomining que puedan ser compartidas y replicadas en Chile y el extranjero, a través de papers, ponencias en congresos y reuniones de bibliotecología y áreas afines.
- Interés profesional por optimizar el uso de los datos que se obtienen desde las distintas fuentes, y ver una metodología para utilizar aquellos logs que tienen datos valiosos y hasta ahora no se han usado en profundidad.,
- Interés profesional por cooperar en el Proyecto Fondecyt 1131109: “Learning analytics: understanding relationships between eLearning and library use, student perceptions of

these tools and learning outcomes" de la Pontificia Universidad Católica de Chile, del que se formó parte como alumna tesista.

- Interés profesional por conocer los patrones de uso multidisciplinarios de nuestros usuarios, con el fin de obtener pautas que permitan diferenciar los servicios y recursos ofrecidos a las distintas unidades académicas.

La finalidad de este trabajo es explorar la relación entre rendimiento académico de los estudiantes y el uso de las colecciones de la biblioteca, junto a un análisis de la multidisciplinariedad de las consultas y su evolución en el tiempo, aplicando técnicas de minería de datos.

Específicamente, por un lado se pretende identificar y explorar perfiles de usuarios cuyos préstamos de material en biblioteca reflejan consultas multidisciplinarias, con una evaluación en el tiempo, y determinar si existe alguna relación con el rendimiento académico.

Por otro lado se busca explorar la relación entre uso de los servicios de la colección de la biblioteca, presencial y electrónica, y el rendimiento de los alumnos, con una evaluación en el tiempo.

Con lo anterior se persigue responder las siguientes preguntas que dirigirán la investigación: ¿Cuál es la relación entre los patrones de uso de los servicios de la biblioteca y el rendimiento académico de los alumnos?, ¿Cuáles son los patrones de uso de los usuarios de la colección presencial de Bibliotecas UC y cuál es el nivel de multidisciplinariedad de sus consultas?, ¿Cuáles son los patrones de uso de los usuarios de la colección electrónica de Bibliotecas UC?, y finalmente ¿Existe una relación entre los patrones encontrados y los resultados de rendimiento académico?

Frente a estas cuestiones de plantearon las siguientes hipótesis:

- ✓ $H_{o:}$ El 50% de las consultas son multidisciplinarias.
  H_1: Menos del 50% de las consultas son multidisciplinarias.

- ✓ $H_{0:}$ No existe diferencia entre el rendimiento académico de los alumnos que usan los servicios de la biblioteca por sobre el promedio.
  $H_{1:}$ Los alumnos con mayor rendimiento son los que usan los servicios de la biblioteca por sobre el promedio.

- ✓ H_0: No existe relación entre las consultas a la colección electrónica y el uso de la biblioteca presencial.
  H_1: Existe relación entre las consultas de la colección electrónica y el uso de la biblioteca presencial.

Para desarrollar la investigación se utilizará el método del estudio de caso, ya que se pretende realizar una investigación empírica de un fenómeno en Bibliotecas UC, que incluye recolección de datos, arreglo de datos y estrategias de análisis de datos para converger en un resultado que se contrastará con las hipótesis planteadas y la literatura investigada.

La estructura de este trabajo comienza con la formulación del problema, luego se plantean las preguntas, objetivos y justificación de la investigación (capítulos 1 al 3).

A continuación se incluye el marco teórico en dónde se recogen las ideas fundamentales de la literatura consultada y los principales conceptos y términos utilizados en el documento. Para luego establecer las hipótesis que serán contrastadas con los resultados (capítulos 4 y 5).

El capítulo siguiente describe los aspectos metodológicos empleados para realizar el estudio, el proceso de recolección de datos y la descripción de los conjuntos de datos con los que se trabajó (capitulo 6).

Posteriormente se incorporan los resultados de los análisis aplicados (Capítulo 7), en primer lugar se incluye los resultados de las consultas a la colección presencial, considerando análisis estadísticos, conjuntos frecuentes de áreas consultadas, entropía de estas consultas, reglas de asociación por área, consultas por área del material y del usuario, consultas por área y notas, clustering de transacciones y notas.

Después se incorporan los análisis realizados para los datos de la colección electrónica, considerando análisis estadístico, consultas por área y carrera del usuario, consultas a la

colección electrónica y uso de la biblioteca, análisis de dominios consultados, conjuntos frecuentes de dominios, reglas de asociación de dominios y clustering de sesiones, duración y notas.

Finalmente se incluyen las conclusiones de ambos análisis (capítulo 8), se realiza la contrastación con las hipótesis y se incorporan sugerencias de aplicación práctica en Bibliotecas UC.

1. Formulación del problema

Diagnóstico de la situación

El rol de las bibliotecas en los ambientes académicos está bastante definido, su importancia es clara, ya que contribuyen al logro de la misión de la institución en la que están inmersas y forman parte del proceso de aprendizaje, enseñanza e investigación, ofreciendo servicios y recursos de información y facilitando a la comunidad académica el descubrimiento, acceso y uso de la información. En el marco de las acreditaciones de las universidades chilenas se contempla un ítem específico respecto a las condiciones que deben cumplirse en cuanto a colecciones e infraestructura que poseen las bibliotecas. En el marco de apoyo a la docencia, cada curso tiene su bibliografía básica que existe en biblioteca, para la investigación buscar en las colecciones de revistas electrónicas de la biblioteca siempre es un buen comienzo, incluso la biblioteca gana popularidad como espacio no sólo de estudio, sino también como de encuentro y distracción.

A pesar de su innegable importancia, la biblioteca debe reinventar su rol en los ámbitos académicos actuales, debe jugar un papel más protagónico en el apoyo a la docencia e investigación. Uno de los grandes desafíos es apoyar el proceso de aprendizaje de los estudiantes, abordando el área de investigación de Learning analytics.

Cabe preguntarse ¿existe una relación entre el logro académico de los estudiantes y su relación con los recursos de la biblioteca?. ¿Cómo lo podemos demostrar?.

En el sistema de Bibliotecas UC se realizan diversos análisis estadísticos para cumplir con los informes y reportes necesarios para la gestión del mismo, como estudios de usuarios, estudios de uso de colecciones y servicios, entre otros, y además para apoyar el proceso de acreditación de diversas unidades de la Universidad. Las limitaciones de estos estudios radican en que responden principalmente a informes requeridos para la gestión, son escasos los recursos y el tiempo para generar investigación en ámbitos distintos, como por ejemplo en experiencia de usuarios, aunque se han desarrollado algunos casos en esta área; estudios comparativos a nivel nacional e internacional, análisis de consultas, entre otros.

Respecto a los estudios de usuarios, los análisis realizados se utilizan principalmente para detectar necesidades de información de los usuarios, programar talleres, generar planes de difusión de servicios, detectar brechas para incluir nuevos servicios, etc.

Dentro de todos estos estudios se ha detectado que no existe conocimiento ni experiencia respecto a la relación entre el uso de los servicios de Bibliotecas UC y el rendimiento académico de los estudiantes, no se ha establecido si hay algún grado de incidencia entre uno y otro.

Respecto a la multidisciplinariedad de las consultas de los usuarios tampoco se han realizado estudios, los que nos permitirían conocer patrones de uso de nuestros usuarios.

La causa probable de esto radica en el desconocimiento que existe sobre cómo realizar los análisis, la dificultad para conseguir los datos de rendimiento académico de los estudiantes y la disponibilidad de tiempo y personal para realizarlas.

Sin estos análisis, se desconoce el rol de Biblioteca como factor influyente en el rendimiento académico de los estudiantes, y no se conocerá la tendencia en la multidisciplinariedad de las consultas de nuestros usuarios.

Por lo tanto será de importancia para la gestión de Bibliotecas UC, y para su posicionamiento dentro de la Universidad, contar con análisis que permitan establecer formalmente, con herramientas y tecnología adecuadas, cuál es la incidencia o relación entre el uso de nuestros servicios y el rendimiento académico de los estudiantes. Asimismo conocer la multidisciplinariedad de las consultas de nuestros usuarios nos permitirá apoyar áreas como desarrollo de colecciones y difusión, dentro de la Biblioteca.

1.1 Preguntas para la investigación

De acuerdo a lo anterior las principales interrogantes que se pretende responder con este estudio son, a nivel general:

¿Cuál es la relación entre los patrones de uso de los servicios de la biblioteca y el rendimiento académico de los alumnos?

Como parte de la sistematización del problema, se espera responder:

> ¿Cuáles son los patrones de uso de los usuarios de la colección presencial de Bibliotecas UC?
>
> ¿Cuál es el nivel de multidisciplinariedad de las consultas a la colección presencial?
>
> ¿Cuáles son los patrones de uso de los usuarios de la colección electrónica de Bibliotecas UC?
>
> ¿Existe una relación entre los patrones encontrados y los resultados de rendimiento académico?

2. Objetivos

Objetivo de la investigación

Explorar la relación entre rendimiento académico de los estudiantes y el uso de las colecciones de la biblioteca.

Objetivo general

Explorar la relación entre rendimiento académico de alumnos de pregrado, uso de las colecciones de biblioteca y multidisciplinariedad de las consultas, aplicando técnicas de minería de datos.

Objetivos específicos

1. Identificar y explorar perfiles de usuarios cuyos préstamos de material de biblioteca reflejan consultas multidisciplinarias, con una evaluación en el tiempo.
2. Explorar relación entre uso de los servicios de la colección de la biblioteca, presencial y electrónica, y rendimiento de los alumnos, con una evaluación en el tiempo.

3. Motivación/Justificación:

En primer lugar el tema de investigación fue abordado principalmente como aporte, en mi calidad de participante como alumna tesista, al proyecto Fondecyt 1131109: “Learning analytics: understanding relationships between eLearning and library use, student perceptions of these tools and learning outcomes” de la Pontificia Universidad Católica de Chile. Cuyos investigadores principales son: Magdalena Jara, Carlos González, Maximiliano Montenegro, Nick Kelly, Elvira Saurina y Rosa Alarcón.

La justificación de este proyecto por una parte fue:

- Las universidades están utilizando cada vez más tecnología para los procesos de enseñanza y aprendizaje, como Sistemas de Manejo del Aprendizaje (LMS) o recursos electrónicos de la biblioteca.
- Investigaciones acerca del proceso de aprendizaje de los alumnos (student learning research) han tendido a observar cómo los estudiantes utilizan estos recursos para su aprendizaje.

Por otra parte:

- Las universidades han avanzado en instalar sistemas electrónicos que van guardando información sobre el uso de estos recursos electrónicos.
- Tendencia en investigación a relacionar la enorme cantidad de datos que acumulan las universidades en estas bases electrónicas con los procesos de aprendizaje de los alumnos: Learning Analytics.

Por lo tanto se hace relevante una investigación sobre cómo los estudiantes en una universidad utilizan estos recursos electrónicos para el aprendizaje y cuál es el impacto en sus resultados de aprendizaje, utilizando bases de datos de uso de los sistemas.

El tema de estudio es nuevo en Bibliotecas UC, no se han realizado anteriormente análisis profundos de los logs de Ezproxy, patrones de uso de los usuarios y su rendimiento académico. Por lo que se considera puede ser un aporte en cuanto a análisis y metodología.

Para lograr los objetivos de estudio, se utiliza técnicas de análisis estadístico y de minería de datos, con las que se pretende establecer una metodología para realizar un análisis de la relación entre rendimiento académico de los alumnos y el uso de los recursos y servicios de la biblioteca, así como un análisis de los patrones de uso de nuestros usuarios, que pueda ser replicable en el tiempo y en distintas instituciones.

Se pretende que esta metodología:

- Sea un apoyo para la gestión de la biblioteca, al tener claridad en la dimensión del uso de los recursos y servicios, tanto presenciales como electrónicos. Además al conocer la relación de la biblioteca con el rendimiento académico de los estudiantes se puede motivar la afluencia de usuarios presenciales y justificar ante las autoridades de la universidad nuevos proyectos de la biblioteca.
- Permita mejorar los servicios ofrecidos, detectando aquellos servicios menos utilizados o con una tendencia a la baja, conociendo los grupos de usuarios que presentan un bajo uso para establecer medidas de difusión entre ellos. Además se puede potenciar aquellos servicios que van en aumento y reenfocar aquellos que se han visto disminuidos.
- Permita detectar necesidades en grupos de usuarios específicos, al conocer sus intereses disciplinarios y sus patrones de consulta, se puede orientar y apoyar el desarrollo de colecciones en la biblioteca, las actividades de desarrollo de competencias informacionales y actividades conjuntas entre las unidades académicas y la biblioteca.

4. Marco teórico y conceptual

4.1 Marco teórico

Las bibliotecas en su actividad diaria producen una gran cantidad de datos, que como dice Marshall Breeding:

> "*Con las herramientas y procesos adecuados, pueden ser explotados para refinar flujos de trabajo, analizar gastos y presupuestos, o informar otras decisiones estratégicas u operacionales*" (Breeding, 2013).

En la medida que crece la diversidad de las colecciones y servicios ofrecidos por las bibliotecas, estas se hacen más complejas generando un desafío para la gestión de los recursos provenientes de distintas fuentes y que yacen en distintas aplicaciones.

Es por esta razón que el área de análisis y datos es una de las características claves de la nueva generación de gestión de bibliotecas (Breeding, 2013)

Muchos productos de ILS (Integrated Library System) incluyen herramientas de reportes, pero estas no son suficientes para proveer el nivel de análisis necesario. Las bibliotecas pueden apoyarse en otros sistemas para realizar análisis más complejos, pero el desafío principal es integrar otras fuentes relevantes que están fuera del ILS.

Frente a esta necesidad de integrar las distintas fuentes de recursos bajo una aplicación que permita gestionarlos surgen las plataformas de servicios bibliotecarios, las que permiten administrar tanto las colecciones digitales como las impresas de la biblioteca, con un fuerte componente para generar reportes y análisis de los datos, que permitiría apoyar la toma de decisiones en la organización (Breeding, 2013).

Algunas bibliotecas han generado sus propias herramientas para realizar minería de datos operacionales, por ejemplo la Universidad de Harvard desarrolla un software de código abierto denominado "Harvard Library Analytics Toolkit", el cual crea visualizaciones de los datos (Breeding, 2013).

Los análisis de datos provenientes de los servicios ofrecidos a través de las aplicaciones en páginas web de las bibliotecas son un componente importante para las organizaciones junto al análisis del diseño del sitio web y experiencia de usuario.

Las analíticas del sitio web de la biblioteca deben ser usadas no sólo para reportar actividad sino también una herramienta para mejorar usabilidad o guiar la toma de decisión respecto a a las herramientas y servicios ofrecidos.

Como las analíticas se vuelven más sofisticadas y se extienden a una amplia área de las actividades de la biblioteca, ellas crean oportunidad para ayudar a las bibliotecas a formar sus colecciones, redistribuir recursos y afinar los flujos de trabajo de las tareas que realizan.

No sólo es importante reflejar la actividad pasada, los datos necesitan ser usados en una forma más predictiva para informar decisiones operacionales e influenciar acciones futuras.

> "*Dado que las bibliotecas usan distintas aplicaciones y servicios que crean inmensas cantidades de datos a través del uso diario, dejar esos datos sin explotar podría representar una desafortunada perdida de oportunidad. Ya sea a través de explotar los reportes y capacidades analíticas disponibles en aquellos sistemas, adquiriendo sistemas especializados de decisión de negocios, o a través de programación local, yo urgiría a las bibliotecas a no permitir que los datos permanezcan en barbecho, si no explotarlos a su potencial completo*." (Breeding, 2013)

Los datos proliferan más rápido de lo que ellos pueden ser capturados, gestionados o almacenados. El valor que se les da a los datos depende del tipo de organización y del uso que se les dé, mientras algunos pasan por un proceso de curación, evaluación de sistemas y generación de políticas, otros no son considerados tan valiosos como para mantenerlos, mientras otros de gran valor no pueden ser mencionados ni citados, ya que forman parte de investigaciones en curso (Wynholds, Wallis, Borgman, Sands & Traweek, 2012)

Dos criterios para medir el potencial analítico de los datos: integridad y ajustes al propósito, son los reconocidos al momento de decidir la curación de los datos. Sin embargo reconocer el potencial de dominio de interés para el reuso de los datos requiere conocimiento de prácticas y necesidades de distintas disciplinas. La evaluación del potencial analítico de los datos será incrementalmente importante para las bibliotecas y repositorios para tomar

decisiones informadas acerca de recolección y curación de datos para ciencias interdisciplinarias (Palmer, Weber & Cragin, 2011).

Encontrar patrones de uso podría permitir encontrar evidencia adicional de dominios potenciales de interés para los datos de la biblioteca (Palmer, Weber & Cragin, 2011).

Es en este punto en donde la minería de datos puede brindar ventajas a las organizaciones para descubrir nueva información que los análisis estadísticos comunes no representan.

Ya en el año 2000, K. Guenther en su "Applying data mining principles to library data collection" indicaba que:

> "*Minería de datos es el proceso de identificar patrones, tendencias y correlaciones a través de grandes cantidades de datos recolectados.*
> *Tal como las industrias las Bibliotecas también desean dividir y analizar sus datos para tomar decisiones basadas en evidencia. Pero las Bibliotecas enfrentan dificultades para recolectarlos, prepararlos, analizarlos e interpretarlos, ya que sus datos están divididos en ambientes físicos y virtuales*."

Los datos recogidos en las Bibliotecas pueden contener información valiosa que puede ser integrada en la gestión estratégica de la biblioteca, y puede ser usada para mejorar la toma de decisiones. Se requiere de una herramienta de análisis automático y descubrimiento para extraer conocimiento útil desde las enormes cantidades de datos de la biblioteca (Chang & Chen, 2006).

Para entender un poco mejor que es la minería de datos, se rescató desde algunos papers sus características y los pasos del proceso.

La minería de datos es un paso en el proceso de descubrimiento del conocimiento, en donde datos de bajo nivel se convierten en conocimiento de alto nivel. El proceso completo de descubrimiento de conocimiento lo presentan Chang & Chen (2006) en la siguiente figura:

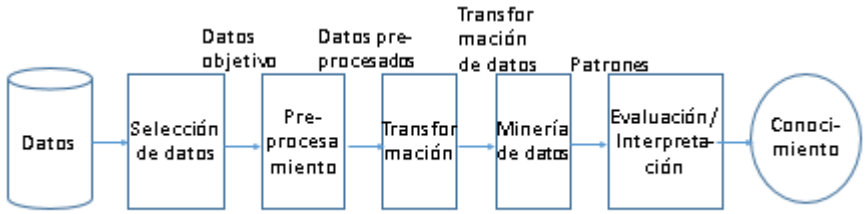

Imagen 4—1: Proceso de descubrimiento de conocimiento

El objetivo de este proceso es ordenar grandes cantidades de datos y descubrir nueva información. La minería de datos proporciona como beneficio convertir el nuevo conocimiento en resultados prácticos.

La minería de datos como se observa en la imagen, es la que permite identificar patrones en los datos los que se convertirán en nuevo conocimiento. El proceso específico de minería de datos lo presentan Chang & Chen (2006) en la siguiente figura:

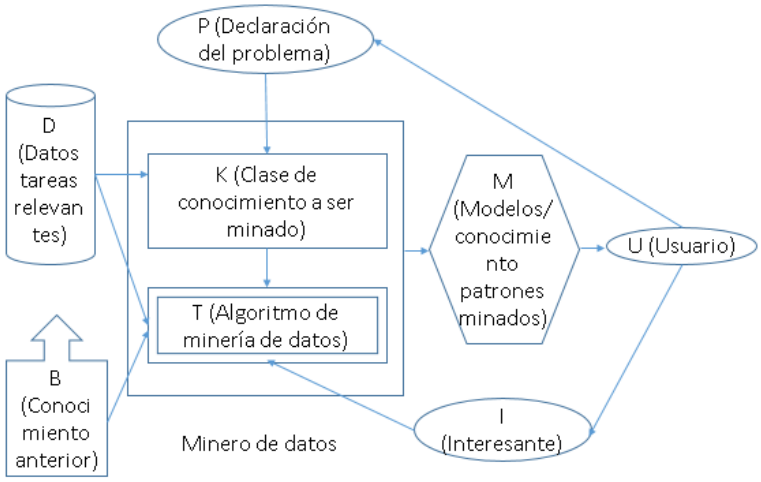

Imagen 4—2: Proceso de minería de datos

Según Chang & Chen (2006), la meta de la minería de datos es extraer información valiosa y nueva desde datos existente, mediante alguna de las siguientes funciones: clasificación, clustering, estimación, predicción, agrupación por afinidad, descripción, etc. La tecnología de minería de datos puede ser dividida en tradicional, como los análisis estadísticos, y la refinada, como la inteligencia artificial, que usa arboles de decisión, redes neuronales, algoritmos genéticos, lógica fuzzy, reglas de inducción, etc.

De acuerdo a Chang & Chen (2006), los modelos de minería de datos más usados son los siguientes:

- Clasificación, clasifica un ítem de dato en una de varias clases de categorías predefinidas.
- Regresión, mapea un ítem de dato en una predicción de variable de valor real.
- Clustering, mapea un ítem de dato en uno de varios clusters, donde los clusters son agrupaciones naturales de ítems de datos basados en métricas de similaridad o modelos de probabilidad de densidad.
- Reglas de generación, extrae reglas de clasificación desde los datos.
- Descubrimiento de reglas de asociación, describe una relación de asociación entre diferentes atributos.
- Sumarización, provee una descripción compacta para un subconjunto de datos.
- Modelación de dependencia, describe dependencias significativas entre variables.
- Análisis de secuencia, modela patrones secuenciales, como análisis de series de tiempo.

Nie & Zhou en su *Study on library administrate decision using data mining* (2012), entrega algunas definiciones respecto al método de clustering, se rescatan algunos aspectos que se considera interesante para los fines de este estudio:

Clúster, es el atributo de un ítem agrupado en clases; agrupa los itemes por la distancia más pequeña o por el nivel de similaridad más alto que se pueda lograr. Clustering es un proceso de aprendizaje no supervisado, lo que lo diferencia de la clasificación, ya que la clasificación necesita conocer anticipadamente cual es la base de las características de los datos, y el objetivo del clustering es encontrar estas características de los datos, por lo tanto, en muchas aplicaciones, el análisis de clúster, se utiliza como un tipo de preproceso de datos adicional al análisis y procesamiento. Como una función de minería de datos, el análisis de clúster puede ser usado cómo una ventaja.

A continuación se entrega una breve descripción de los principales métodos de clustering (Nie & Zhou, 2012), algunos de ellos son:

- Algoritmo K-Means, busca cada clase de objeto con la media declarada.
- K-center en el algoritmo, algoritmos, busca un centro de clustering cercano a los distintos objetos.

- Clustering jerárquico, hay dos tipos el aglomerativo, también conocido como el método desde abajo hacia arriba, considera cada objeto un grupo separado, y entonces mezcla el objeto o similares, hasta que todas las combinaciones alcancen una terminación, es decir sean agrupados en una clase. El otro es el clustering jerárquico, también conocido como método de arriba hacia abajo, comienza con todos los objetos colocados en una clase, en cada paso de la iteración una clase es dividida en clases más pequeñas, hasta que cada objeto quede en una clase separada.
- Método basado en la densidad, la mayoría de los métodos de división está basado en la distancia entre los objetos del clúster. Para el método basado en la densidad de clustering la idea es: cercano a la densidad de un objeto o área (el número de puntos de datos) más que un cierto límite, continuará generando el clustering. Cada clase corresponde a cada uno de los puntos dentro de un rango dado, se debe considerar métodos para filtrar ruidos o datos fuera de rango.

Esta metodología, utilizada ampliamente en áreas de las ciencias de la ingeniería también ha llegado a las ciencias de la información, y específicamente a las bibliotecas. Chang & Chen (2006), describen los siete pasos en el proceso de minería de datos para bibliotecas:

1. Establecer metas de minería. Decidiendo cuales son los resultados deseados.
2. Seleccionar los datos. Decidir cuales datos son útiles, que atributos son valiosos para considerar, y cuál debe ser el tamaño de la muestra.
3. Pre-proceso de los datos. Filtrar datos que producen ruido, erróneos o irrelevantes y manejar datos perdidos.
4. Transformar los datos. Donde sea posible, reducir el número de atributos de datos o extraer nuevos desde atributos de datos existentes. Combinar tablas de datos y proyectar los datos en espacios de trabajo, esto es, tablas que representan el nivel de abstracción óptima para el problema de interés.
5. Almacenar los datos. Integrar y almacenar datos en un solo sitio bajo un esquema unificado.
6. Minería de datos. Realizar las funciones y algoritmos de minería de datos apropiados de acuerdo a metas de minería. Los analistas primero construyen cubos de datos para proveer vistas multidimensionales de los datos. Luego ejecutan

minería analítica en línea usando la estructura de cubos de datos multidimensionales para descubrimiento de conocimiento.

7. Evaluar resultados de minería. Ejecutar varias operaciones tal como filtrado de conocimiento desde la salida, analizando la utilidad del conocimiento extraído, y presentando los resultados al usuario para retroalimentación. La retroalimentación de este paso puede producir cambios en los pasos anteriores.

En diversos lugares existen experiencias de minería de datos en bibliotecas, como la presentada por Gibson (2001) en su "*Data mining analysis of digital library database usage patterns as a tool facilitating efficient user navigation".* Aquí detalla la experiencia de desarrollo de un método de recomendaciones para una biblioteca digital clínica en línea, usando métodos diseñados localmente, datos obtenidos desde logs de un servidor, previamente limpiados, con el fin de facilitar la minería de datos de los patrones de navegación de los usuarios.

Nicholson en su "*The bibliomining process: Data warehousing and data mining for library decision making"* (2003), nos presenta un término que representa su experiencia de aplicación de minería de datos en biblioteca: "Bibliomining", que corresponde a la aplicación de minería de datos y herramientas bibliométricas a datos producidos desde servicios de Bibliotecas.

Este término fué usado por primera vez por Nicholson y Stanton en su *"Gaining strategic advantage through bibliomining: data mining for management decisions in corporate, special, digital, and traditional libraries" (2003).*

El proceso de bibliomining es un proceso cíclico (Nicholson, 2003), que consiste de los siguientes pasos:

- determinar áreas de foco
- Identificar fuentes de datos internas y externas
- Recoger, limpiar y anonimizar los datos en un data warehouse.
- Seleccionar herramientas de análisis apropiadas.
- Descubrir patrones a través de minería de datos y creación de reportes con herramientas analíticas tradicionales; y
- Analizar e implementar los resultados.

Otra experiencia interesante de aplicación de minería de datos en bibliotecas es la que contribuye a la gestión de alto nivel (Nie & Zhou 2012). Las bibliotecas modernas gestionan toda clase de datos producidos por sus sistemas automatizados, la administración tradicional ya no se adapta al desarrollo de las bibliotecas modernas, se requiere nuevas herramientas y tecnología, la minería de datos responde a los nuevos desarrollos. Usar esta tecnología permitirá obtener una toma de decisiones científica que entrega potentes garantías.

Entre los objetivos de aplicar minería de datos a la gestión de la biblioteca, se puede considerar: entender el propósito de los accesos y tendencias de los usuarios de la biblioteca, entender los intereses y demandas de los lectores, mejorar la calidad del servicio y satisfacer la demanda de los lectores, mejorar la satisfacción de los lectores, dar a los lectores un mejor servicio, mejorar la toma de decisiones en la biblioteca.

Nie & Zhou en su *Study on library administrate decision using data mining* (2012), aportan las siguientes recomendaciones:

- La biblioteca en el proceso de minería de datos. La minería de datos en la biblioteca tiene una estructura dividida en tres niveles: nivel de datos, nivel de tecnología y nivel de equipo de gestión. El nivel de datos tiene como trabajo principal recoger datos de la colección, préstamos y servicios de la biblioteca, redes por las que se consulta la colección en línea de la biblioteca, recursos humanos y recursos de hardware, información histórica de usuarios, de préstamos y de la biblioteca. El nivel de tecnología, se enfoca en usar análisis de clúster y de correlación, y otros modelos, trabaja con herramientas de minería de datos. El nivel de equipo de gestión, su principal tarea es la toma de decisiones, utiliza el análisis realizado por el nivel de tecnología en la toma científica de decisiones.
- La organización de la colección de datos de la biblioteca. De acuerdo a las necesidades de la minería de datos se utilizan diversos tipos de datos, como diseño de cuestionarios y minería de datos, recolección y ordenamiento de datos de la biblioteca, datos de la red general de la biblioteca y del sistema integrado. Por lo que se requiere mantener identificados los datos relevantes para la gestión.

Otra experiencia la presentan Hinze, McKay & Vanderschnantz en su *Book Selection Behavior in the Physical Library: Implications for eBook Collections* (2012). Es un estudio sobre cómo la gente selecciona libros físicos desde estanterías de bibliotecas académicas. Se utilizan percepciones obtenidas a través de observaciones y entrevistas respecto a conductas de selección de libros, el objetivo es hacer sugerencias para el diseño de bibliotecas digitales de libros electrónicos para facilitar una mejor conducta de selección de libros.

Kawase & Herder en su *Classification of User Interest Patterns Using a Virtual Folksonomy* (2011), nos presenta su experiencia con el estudio de asociaciones de patrones de uso con clustering. El interés de los usuarios en ciertos tópicos y recursos es recurrente y sigue ciertos patrones, los métodos tradicionales para predecir patrones recurrentes están basados en ranqueo y modelos asociativos. En este estudio se identifican varios patrones canónicos realizando clustering de palabras claves relacionadas con los recursos visitados, utilizando un gran repositorio de datos de uso del web. Las palabras claves son derivadas desde una folksonomia virtual de etiquetas asignadas a los recursos usando un sistema de etiquetado colaborativo.

Otro tema asociado a bibliotecas, específicamente a las búsquedas de información, es el de los topics models o modelos de tópicos, Newman, Noh, Talley, Karimi & Baldwin nos presentan su experiencia de trabajo en su *Evaluating Topic Models for Digital Libraries* (2010). Los módelos de tópicos pueden tener un enorme impacto en mejorar las formas en que el usuario encuentra y descubre contenido en bibliotecas digitales e interfaces de búsqueda, a través de su disponibilidad para aprender y aplicar automáticamente etiquetas de materia a cada uno de los itemes en una colección, y su capacidad para crear dinámicamente colecciones virtuales en el aire.

Las etiquetas de contenidos han permitido realizar múltiples estudios respecto a la forma en que buscan nuestros usuarios y los sistemas que incluyen etiquetado social van en aumento. Sun & Lin en su *Topical community detection from mining user tagging behavior and interest* (2013), proponen un modelo generativo probabilístico para detector comunidades latentes de tópicos entre los usuarios. Las etiquetas sociales y contenidos de los recursos son controlados para modelar los intereses de los usuarios en dos formas similares y correlacionadas. La meta principal es capturar las conductas e intereses de

etiquetado de los usuarios y descubrir los tópicos emergentes en una estructura de comunidad. Las comunidades son grupos de usuarios con interacciones sociales frecuentes y tópicos similares de interés, que podrían tener importantes implicaciones en investigaciones para servicios de información personalizados. Los resultados experimentales en dos conjuntos de datos reales con etiquetado social de diferentes géneros, muestran que el modelo propuesto precisa los modelos de interés de los usuarios y detecta con alta calidad comunidades de tópicos significativas.

De la aplicación de la minería de datos en documentos, se acuña un nuevo concepto el de "Minería de textos", que Chen, Y., Liu, Y., & Ho, W., en su artículo "*A text mining approach to assist the general public in the retrieval of legal documents*", definen como:

> "*El proceso de descubrimiento de información fundamental y característica desde datos textuales. La minería de textos es un campo interdisciplinario de varias tecnologías, incluyendo bases de datos, minería de datos, recuperación de información, lingüística, estadística y otras*" (Chen, Liu & Ho, 2013)

Existen también experiencias de aplicación de minería de datos relacionados con la biblioteca como recurso de aprendizaje, Palmer presenta un ejemplo en su *Web usage mining: Application to an online educational digital library service* (2012). Este estudio involucra la minería de datos educacional, bibliotecas digitales educacionales y el análisis de conductas de los profesores mientras crean recursos de aprendizaje en línea en el Sistema Instructional Architect (IA). Una experiencia con la misma herramienta la relatan Xu & Recker en su *Teaching analytics: A clustering and triangulation study of digital library user data* (2012), exploran la conducta de los profesores al utilizar la biblioteca digital a través de clustering y conjuntos frecuentes. Sus resultados presentan que los años de experiencia en enseñanza y el confort con la tecnología estaban relacionados con la efectividad de los profesores al usar el IA.

Respecto a la búsqueda de la relación entre el uso de la biblioteca y el rendimiento de los alumnos, existen diversas experiencias con estudios enfocados en demostrar el valor de la biblioteca para la universidad en un contexto de limitaciones económicas, en donde se requiere competir por los recursos con otras unidades de la misma universidad. En ellos se han encontrado asociaciones positivas entre los resultados académicos y el uso de la

biblioteca (Cox & Jantti, 2012; Goodall & Pattern, 2011; Soria, Fransen, & Nackerud, 2013; Stone & Ramsden, 2013).

En educación existen diversas experiencias en cuanto a la medición del aprendizaje, frente a la dificultad de medir cuanto alguien ha aprendido los investigadores buscan medidas efectivas de aprendizaje, algunos mediante la utilización de minería de datos en resúmenes manuscritos producidos por los participantes (Wilson & Wilson, 2013).
Además la minería de datos se aplica en diversas disciplinas, como Derecho y Salud, como se señala en las experiencias descritas en los artículos de Chen, Liu & Ho (2013) y de Ku, Chiu, Zhang, Chen & Su (2014).

Otras experiencias de utilización de la minería de datos que fueron consultadas son: análisis de búsquedas de imágenes en la web (Choi, 2013), la efectividad de la recuperación de consultas traducidas por maquinas (Dolamic & Savoy, 2010), ranqueo de resultados (Deng, Han, Lyu & King, 2012), ranqueo y modelación de lenguaje (Lin, Lin, Xu & Sun, 2013), ranqueo en tiempo real (Liu & Turtle, 2013), modelación de subtópicos (Zheng, Wang, Fang & Cheng, 2011), organización del conocimiento (Rodríguez-Bárcenas & López-Huertas, 2013), indexación de semántica latente (Zhu & Allen, 2013), semántica (Afiontzi, Kazadeis, Papachristopoulos, Sfakakis, Tsakonas & Papatheodorou, 2013), y visualización (Buchel, 2011).

Con este marco nos damos cuenta de que existen aplicaciones de la minería de datos en diversas disciplinas, y en la relación que nos interesa para este estudio: educación y bibliotecas, específicamente el análisis de la biblioteca como factor que contribuye al aprendizaje.

4.2 Marco conceptual

Los principales conceptos utilizados durante la investigación son los siguientes:

Aleph, es el sistema integrado que soporta los principales servicios de Bibliotecas UC.

Áreas OCDE, son seis áreas que categorizan el conocimiento en forma general, generadas por la OCDE, en Chile CONICYT los utiliza para identificar las áreas de los proyectos de investigación. Para efectos de este análisis se utilizan sólo cinco de las áreas, considerando el área de Agricultura dentro de Ingeniería y Tecnología: Cs. Naturales, Cs. Sociales, Humanidades, Ingeniería y Tecnología, Cs. Médicas y de la Salud. Se asignaron a los materiales basándose en las áreas temáticas (Ver Anexo 3).

Áreas Temáticas, es un número que identifica en forma general la disciplina que trata el material, se genera a partir del número de pedido, que está basado en la Clasificación Decimal Dewey (Ver Anexo 3).

Clustering o algoritmo de agrupamiento, según la Wikipedia se define brevemente como:

> *"un procedimiento de agrupación de una serie de vectores de acuerdo con un criterio. Esos criterios son por lo general distancia o similitud. La cercanía se define en términos de una determinada función de distancia, como la euclidiana, aunque existen otras más robustas o que permiten extenderla a variables discretas. La medida más utilizada para medir la similitud entre los casos es la matriz de correlación entre los nxn casos. Sin embargo, también existen muchos algoritmos que se basan en la maximización de una propiedad estadística llamada verosimilitud*.

Colección electrónica, se refiere a la colección de la Biblioteca compuesta por recursos que se pueden consultar en línea, sin necesidad de asistir a la biblioteca, tales como libros, revistas, tesis, bases de datos y material digitalizado.

Colección presencial o física, se refiere a la colección de la Biblioteca compuesta por material impreso y audiovisual, que requiere la presencia del usuario en biblioteca para retirarlo en préstamo. No se considera los espacios de la biblioteca (salas y cubículos).

Confianza, es un indicador que señala la certeza que se tiene de que uno o varios itemes aparezcan relacionados junto a otro u otros dentro de un conjunto de datos.

Conjuntos frecuentes, o frequent itemsets, son grupos de itemes que a menudo aparecen juntos en los datos recolectados. Son la base para determinar las reglas de asociación existentes entre los itemes.

Consultas, cantidad de solicitudes de material en la colección presencial o física de la biblioteca.

DBSCAN, es un operador para realizar agrupaciones o clustering, según Wikipedia:

> *"density-based spatial clustering of applications with noise o agrupamiento espacial basado en densidad de aplicaciones con ruido, propuesto por Martin Ester, Hans-Peter Kriegel, Jörg Sander y Xiaowei Xu en 1996. Es un algoritmo basado en densidad porque encuentra un número de grupos (clusters) comenzando por una estimación de la distribución de densidad de los nodos correspondientes. DBSCAN es uno de los algoritmos de agrupamiento más usados y citados en la literatura científica"*.

Dominios, se refiere a los sitios consultados por los usuarios, específicamente son la raíz de las URLs consultadas.

Entropía, es una medida de la incertidumbre o un indicador que señala el nivel de homogeneidad/heterogeneidad que existe en el comportamiento de una variable en el tiempo.

EZProxy, es un servidor proxy web que permite el acceso remoto desde fuera de la red UC a la colección electrónica de la biblioteca, reconociendo al usuario como miembro de la comunidad UC a través de la autenticación con su clave y contraseña de correo institucional.

K-means, es un operador para realizar agrupaciones o clustering, según la Wikipedia es:

> *"K -means es un método de agrupamiento, que tiene como objetivo la partición de un conjunto de n observaciones en k grupos en el que cada observación pertenece al grupo más cercano a la media. Es un método utilizado en minería de datos".*

Learning Analytics, analíticas de aprendizaje, según la Wikipedia es:

> *"La medición, recopilación, análisis y presentación de datos sobre los estudiantes, sus contextos y las interacciones que allí se generan, con el fin de comprender el proceso de aprendizaje que se está desarrollando y optimizar los entornos en los que se produce".*

Minería de datos, o exploración de datos, el proceso de extraer modelos útiles desde los datos. Algunas veces un modelo puede ser un resumen de los datos o el conjunto de las características más extremas de los datos (Leskovec, Rajaraman & Ullman, 2014).

Según la Wikipedia es:

> *"La etapa de análisis de "Knowledge Discovery in Databases", o KDD, es un campo de las ciencias de la computación referido al proceso que intenta descubrir patrones en grandes volúmenes de conjuntos de datos. Utiliza los métodos de la inteligencia artificial, aprendizaje automático, estadística y sistemas de bases de datos. El objetivo general del proceso de minería de datos consiste en extraer información de un conjunto de datos y transformarla en una estructura comprensible para su uso posterior".*

Reglas de asociación, en este estudio se considera como la relación entre dos valores que tienden a aparecer juntos en un conjunto de datos. según la Wikipedia:

> *"En minería de datos y aprendizaje automático, las reglas de asociación se utilizan para descubrir hechos que ocurren en común dentro de un determinado conjunto de datos".*

Sesiones, corresponden a los accesos a la colección electrónica de Bibliotecas UC, desde fuera de la red UC vía EZProxy o dentro de la Red UC.

Soporte, se refiere al indicador de frecuencia de uno o varios ítemes en un conjunto de datos.

Tipos de material, es un código que identifica la categoría del material de acuerdo a sus características físicas. Para este análisis, sólo se han considerado los materiales bibliográficos, no aquellos de espacio físico ni equipos de la biblioteca. Por ejemplo: libros, tesis, Blue Ray, entre otros.

Tipos de usuarios, es un código numérico que permite categorizar los usuarios de la biblioteca, de acuerdo a su rol dentro de la universidad, por ejemplo, alumnos (egresados, regulares, internos, postdoctorados), docentes (activos, retirados, eméritos), investigador bibliográfico, administrativos (activos y retirados), instituciones y proyectos. (Ver Anexo 2).

Transacciones, para efecto de este análisis, se refiere a la acción de consultar un material o servicio de la Biblioteca. Para la primera parte se considera sólo transacciones de préstamos y renovaciones, es decir, el evento que se registra cuando un usuario solicita en préstamo un material de la biblioteca y cuando lo renueva desde su cuenta. Para la

colección electrónica se considera como una transacción las consultas, sesiones o conexiones que realiza el usuario a los recursos.

Unidad Académica, es un código numérico que identifica la Escuela o Facultad a la que pertenece el usuario.

Usuario: Se refiere a quién hace usos de los servicios de la biblioteca, tanto presenciales como electrónicos.

Zscore de notas, es un cálculo que promedia las notas obtenidas dentro de cada carrera, considerando la desviación estándar de ese grupo específico, con el fin de poder hacer comparaciones equitativas entre las notas de las distintas carreras.

5. Hipótesis

Las hipótesis nulas y las alternativas planteadas en esta investigación, considerando el marco conceptual y la experiencia en el área bibliotecológica, son las siguientes:

- ✓ $H_{o:}$ El 50% de las consultas son multidisciplinarias.
  H_1: Menos del 50% de las consultas son multidisciplinarias.

- ✓ $H_{o:}$ No existe diferencia entre el rendimiento académico de los alumnos que usan los servicios de la biblioteca por sobre o bajo el promedio.
  $H_{1:}$ Los alumnos con mayor rendimiento son los que usan los servicios de la biblioteca por sobre el promedio.

  En el estudio de esta última hipótesis se consideró interesante estudiar la relación entre el uso de la colección presencial y la electrónica, por lo que surgió una hipótesis lateral:

  - ✓ H_0: No existe relación entre las consultas a la colección electrónica y el uso de la biblioteca presencial.
    H_1: Existe relación entre las consultas a la colección electrónica y el uso de la biblioteca presencial

6. Aspectos metodológicos

6.1 Metodología

Se utilizará el método del estudio de caso único (explicativo o descriptivo) ya que se pretende realizar una investigación empírica de un fenómeno, que incluye recolección de datos y estrategias de análisis de datos para converger en un resultado que se contrastará con las hipótesis planteadas y la literatura investigada.

Se considerará el caso único de la relación entre rendimiento académico de alumnos de pregrado de la Pontificia Universidad Católica de Chile, uso de las colecciones de Bibliotecas UC e multidisciplinariedad de las consultas, aplicando técnicas de minería de datos.

Luego del planteamiento del problema, se revisó literatura nacional y extranjera en búsqueda de experiencias similares, y se formularon las proposiciones.

La recolección de datos se realiza desde los registros de bases de datos de la UC, tanto de Bibliotecas UC como de Registros Académicos, específicamente los datos de transacciones presenciales, datos de itemes consultados durante los años 2010-2014, sesiones de consultas de la colección de recursos electrónicos y la información de rendimiento académico y unidades académicas de los alumnos para los años 2012-2014. Las fuentes de recolección de los datos utilizados son explicadas en detalle en el punto 6.3 y los conjuntos de datos o matrices que se obtuvieron son descritos en detalle en el punto 6.4.

Todo el conjunto de datos utilizado junto a la literatura consultada se considera la fuente de evidencia de la investigación, con la que se realizará la triangulación de la información.

Respecto a los aspectos éticos en el uso de los datos privados de los usuarios, para resguardar la confidencialidad de estos se utiliza los datos de identificación de los usuarios encriptados u ocultos.

Para el análisis global y análisis profundo de los datos obtenidos se trabaja con las matrices obtenidas y se realiza un análisis estadístico de los datos, utilizando distintas variables de estudio que se combinan en los análisis, estas variables fueron seleccionadas considerando su relevancia para dar respuesta a las preguntas de investigación y las hipótesis formuladas:

- ✓ Cantidad de préstamos,
- ✓ Cantidad de sesiones de consultas de colección electrónica,
- ✓ Duración de las sesiones,
- ✓ Tipo de usuario
- ✓ Carrera,
- ✓ Año de la carrera,
- ✓ Año de la transacción/sesión,
- ✓ Área temática del material consultado
- ✓ Área del usuario
- ✓ Dominios de los recursos electrónicos consultados
- ✓ Notas del alumno, expresadas en Zscore por carrera.

Para cada análisis se generan tablas y se grafican los resultados.

Además se creó nuevos set de datos con algunas de estas variables a los que se aplicaron técnicas de minería de datos como:

- ✓ Conjuntos frecuentes, se consideró las áreas temáticas del material y los dominios de los recursos electrónicos consultados.
- ✓ Modelo de Reglas de asociación, se consideró las áreas temáticas del material y los dominios de los recursos electrónicos consultados.
- ✓ Modelos de Clustering, se consideró las notas en Zscore, la cantidad de préstamos, la cantidad de consultas a la colección electrónica, y la duración de estas consultas. Todo esto considerando los años en que se llevaron a cabo.

Finalmente se recopila medidas de desempeño de los modelos utilizados.

Luego se procede a realizar el análisis de los resultados, siguiendo su evolución en el tiempo.

La estrategia seleccionada para realizar este análisis es la contrastación con las proposiciones teóricas o hipótesis. Además se intentará utilizar la técnica de elaborar una explicación del caso.

No se realizará comparación de resultados con otras instituciones nacionales o internacionales similares ya que los datos utilizados no son de libre disponibilidad, las instituciones los mantienen a resguardo por ser considerados datos privados de las personas de cada Universidad.

Se podrá contrastar, en algunos casos, los resultados con investigaciones similares realizadas en otras instituciones extranjeras, que se encuentran disponibles en publicaciones periódicas o en Internet.

Finalmente se establecen las conclusiones e implicancias de la investigación.

6.2 Recolección de Datos

Se trabajó con distintas matrices de datos obtenidos desde las bases de datos de la Universidad y de Bibliotecas UC.

Para la obtención de los datos de la matriz de Consultas de la colección presencial 2010-2014 se trabajó con el software Infomaker, que trabaja sobre la base de datos de Aleph, el software de Bibliotecas UC.

Durante marzo y abril, se realizaron varias consultas SQL hasta llegar a la que permitía obtener los datos deseados, dejando sólo los datos de transacciones de material bibliográfico de usuarios personales.

Se generó una consulta por cada año (Ver anexo 1), se revisaron los datos obtenidos, se filtraron y ordenaron generando un archivo Excel para cada año.

En cada archivo Excel se trabajó los datos con tablas dinámicas para obtener listados por área.

Para la Información de Rendimiento académico y transacciones de los alumnos 2012-2014 se descargó la matriz generada por el proyecto Fondecyt, desde su base de datos en PosgreSQL. Esta fue filtrada, dejando las columnas relativas a rendimiento académico, transacciones y año.

Para generar la Matriz de rendimiento académico, áreas del material, carrera y zscore del alumno 2012-2014, se realizó un pareo de datos a través de los ID de los usuarios para unir las transacciones por área temática obtenidas en la primera tabla (Consultas de la colección impresa 2010-2014), con la información de carrera, rendimiento y el Zscore por carrera obtenidas de la Matriz de rendimiento académico y transacciones 2012-2014.

La Matriz de Zscore y EzProxy, sesiones y consultas, 2012-2014, es producto de la Matriz de rendimiento académico y transacciones 2012-2014, se generó una tabla dinámica en donde se dejó sólo las columnas con la información de Zscore, sesiones y duración de Ezproxy. Se eliminaron aquellas filas en que había datos con error.

6.3 Descripción de los conjuntos de datos

a) Consultas de la colección presencial 2010-2014

Este archivo contiene los datos de transacciones para los años 2010 al 2014, realizando una consulta SQL sobre las tablas de datos de Aleph se obtuvo un listado con las siguientes cantidades de filas para cada año:

Tabla 6-1: Cantidad de Consultas de la colección presencial 2010-2014

	Total transacciones	Total usuarios
2010	583885	23318
2011	550534	22677
2012	527733	22799
2013	502431	22723
2014	442748	23218
Total	1474092	114735

El archivo contiene los siguientes atributos o variables para cada fila:

Tabla 6-2: Matriz de Consultas de la colección presencial 2010-2014

z35_event_date	z13_title	z30_inven	z30_call_no	z30_material	z35_id	z35_bor_stat	z35_bor_typ	general	OCDE
20110404	Complexitie	1	001 C737o 2002	LIB	)732	3	91	001	Cs. Sociales
20110512	El desafío de	1	001 J82d.E 2000	LIB	7763	7	57	001	Cs. Sociales
20110113	The postmoc	1	001.01 L991c.I 1984	LIB	9768	7	20	001	Cs. Sociales
20110629	Reason and t	1	001.01 S529r 1984	LIB	3392	3	12	001	Cs. Sociales

1. Fecha de transacción: código numérico de largo 8, con formato Año Mes Día (AAAAMMDD)
2. Título: Cadena de caracteres alfanumérico de largo variable, que identifica la obra.
3. Área temática: código numérico de largo variable, que identifica el área temática del item desde el número de clasificación que tiene asignado el título.
4. Número de pedido: código alfanumérico de largo variable, que incluye número de clasificación, número de autor, año, volumen y/o copia.
5. Tipo de material: código alfabético de largo 5, que identifica el formato del material.
6. Id de usuario: Código alfanumérico que identifica en forma única a los usuarios, principalmente es el RUT.
7. Tipo de usuario: código numérico de largo 2, que identifica la categoría del usuario en la universidad.
8. Unidad académica: código numérico de largo 2, que identifica a la unidad a la que pertenece el usuario.
9. Área OCDE: Cadena de caracteres de largo variable, que identifica el tema de los ítems a partir del área temática. Son cinco valores: Cs. Médicas y de la Salud, Cs. Naturales, Cs. Sociales, Humanidades, Ingeniería y Tecnología.

b) Matriz de rendimiento académico y transacciones 2012-2014 (proyecto Fondecyt)

Esta matriz proviene de la base de datos del proyecto Fondecyt Learning Analytics, contiene 107398 filas para 28756 usuarios. Con datos de los años 2012 a 2014.

Tabla 6-3: Matriz de rendimiento académico y transacciones 2012-2014

ID	nom_carrera	area_compa	ano_ingreso	ano_carrera	prom_notas	lib_digital_c	lib_space_cc	ezproxy_blo	ezproxy_dur	ezproxy_nur	ezproxy_nur	ano	semestre	zscore_carre	zscore_area
023	LICENCIATUF	Humanidade	2012	1	4.46666667	0	0	0	0	0	0	2012	22	-0.887439	-1.220707
165	ARQUITECTU	Ciencias Soci	2012	1	6.26666667	0	0	1	239	1	25	2012	22	1.517359	1.395195
456	ARQUITECTU	Ciencias Soci	2012	3	4.42	0	0	\N	\N	0	\N	2014	21	-1.352202	-0.957145
456	ARQUITECTU	Ciencias Soci	2012	2	4.4	0	0	\N	\N	0	\N	2013	22	-1.174627	-1.028556
456	ARQUITECTU	Ciencias Soci	2012	2	5.325	0	0	0	0	0	0	2013	21	-0.000194	0.194253
456	ARQUITECTU	Ciencias Soci	2012	1	5.46	0	0	0	0	0	0	2012	22	0.244781	0.302242
456	ARQUITECTU	Ciencias Soci	2012	3	4.95	0	0	\N	\N	0	\N	2014	22	-0.570998	-0.4064
520	ACTUACION	Humanidade	2012	2	5.7	0	0	\N	\N	0	\N	2013	22	0.455721	0.401318

No se utilizó la matriz original completa del proyecto, se seleccionaron sólo las siguientes variables:

1. ID: Corresponde a la encriptación del número de identificación de los estudiantes
2. nom_carrera: nombre de la carrera
3. area_compacta: es el área temática de la carrera de acuerdo a las áreas del conocimiento que establece la OCDE. Para efectos de este estudio se dejaron sólo 5 áreas: Cs. Médicas y de la Salud, Cs. Naturales, Cs. Sociales, Humanidades, Ingeniería y Tecnología.
4. ano_ingreso: Año de la primera matricula del alumno en la carrera.
5. ano_carrera: Cálculo del año que cursa el alumno, de acuerdo al año de ingreso.
6. prom_notas: Calculo del promedio de notas anual.
7. lib_print_count: Cantidad de préstamos de material impreso.
8. lib_digital_count: Cantidad de préstamos de material no impreso.
9. lib_space_count : Cantidad de préstamo de espacio en biblioteca.
10. ezproxy_block_ten_minutes: Consultas realizadas en bloques de 10 minutos a través de EZProxy.
11. ezproxy_duration_avg: Promedio de duración de las consultas a través de EZProxy.
12. ezproxy_number_sessions: Cantidad de sesiones del estudiante a través de EZProxy.
13. ezproxy_number_lines_avg: Promedio de número de líneas de consultas a través de EZProxy.
14. Ano: Año de estudio al que corresponden las transacciones.
15. Semestre: Semestre del año de estudio al que corresponden las transacciones.
16. zscore_carrera: Zscore del alumno calculado por carrera.
17. zscore_area: Zscore del alumno calculado por área de la carrera.

c) Matriz de rendimiento académico, áreas del material, carrera y zscore del alumno 2012-2014.

Esta matriz es producto del cruce de las dos tablas anteriores, Consultas de la colección presencial 2010-2014 y Matriz de rendimiento académico y transacciones 2012-2014, por ID de usuario, con el fin de relacionar la información de rendimiento con la de las transacciones en biblioteca y áreas consultadas de los alumnos. Para lograr el cruce de las tablas se desencriptó el ID del archivo Matriz de rendimiento académico y transacciones 2012-2014.

La cantidad de filas por año del archivo que resultó fueron las siguientes:

2012 - 14933 usuarios

2013 - 15017 usuarios

2014 - 16419 usuarios

Tabla 6-4: Matriz de rendimiento académico, áreas del material, carrera y zscore del alumno 2012-2014.

ID1	Carrera	area us	Transaccione	zscore_carrera	Cs. Médicas	Cs. Naturale:	Cs. Sociales	Humanidade	Ingeniería y Tecnología
461	TEOLOGIA	Humanidades	20	0.950789	0	0	0	20	0
497	TEOLOGIA	Humanidades	132	0.343693	0	0	120	12	0
561	ENFERMERIA	Ciencias Médicas y	35	2.293164	0	1	0	34	0
142	TEOLOGIA	Humanidades	93	0.404402	0	1	2	90	0
046	MATEMATICA	Ciencia Naturales	23	0	0	23	0	0	0
753	DERECHO	Ciencias Sociales	1	0.091094	0	0	1	0	0
505	TEOLOGIA	Humanidades	22	0.270841	0	0	0	22	0
274	TEOLOGIA	Humanidades	21	0.70795	0	0	0	21	0
510	TEOLOGIA	Humanidades	2	0.950789	0	0	0	2	0

1. ID1: Número de identificación del estudiante.
2. Carrera: Nombre de la carrera del estudiante.
3. area_us: Es el área temática de la carrera de acuerdo a las áreas del conocimiento que establece la OCDE. Para efectos de este estudio se dejaron sólo 5 áreas: Cs. Médicas y de la Salud, Cs. Naturales, Cs. Sociales, Humanidades, Ingeniería y Tecnología.
4. Transacciones: Suma de transacciones de préstamos impresos, digital y espacio.
5. Zscore_carrera: Zscore del alumno calculado por carrera.
6. Cs. Médicas: Área del material de biblioteca que consultó el estudiante.
7. Cs. Sociales: Área del material de biblioteca que consultó el estudiante.
8. Humanidades: Área del material de biblioteca que consultó el estudiante.
9. Ingeniería y Tecnología: Área del material de biblioteca que consultó el estudiante.

d) Matriz de Zscore y EzProxy, sesiones y consultas, 2012-2014

Esta matriz es un subconjunto de la “Matriz de rendimiento académico y transacciones 2012-2014”, en donde se separó las columnas de promedio de Zscore, las sesiones de EZproxy y su duración, por año de estudio, generando tres listados sobre los cuales se generaron posteriormente los clúster.

La cantidad de filas, correspondientes a estudiantes, por año fue la siguiente:

2012 – 18206 alumnos

2013 – 21343 alumnos

2014 - 21909 alumnos

Tabla 6-5: Matriz de Zscore y EzProxy, sesiones y consultas, 2012-2014

Etiquetas de fila	Promedio de zscore_carrera2	Suma de ezproxy_number_sessions	Suma de ezproxy_duration_avg
456	-0,96	0	0,00
368	-0,04	0	0,00
055	0,35	0	0,00
800	0,30	0	0,00
801	-0,80	0	0,00
802	-0,66	0	0,00
810	-0,47	0	0,00
160	1,04	0	0,00
270	-2,04	27	60,67
189	-0,14	0	0,00
461	0,44	0	0,00
541	0,80	2012	577,61
568	1,09	0	0,00

2012_EP_zscore | 2013_EP_Zscore | **2014_EP_Zscore**

1. Etiquetas de fila: corresponde al ID de los estudiantes
2. Promedio de zscore_carrera2: Zscore del alumno calculado por carrera.
3. Suma de ezproxy_number_sessions: Cantidad de sesiones del estudiante a través de EZProxy.
4. Suma de ezproxy_duration_avg: Promedio de duración de las consultas a través de EZProxy.

7. Resultados

7.1 Consultas de la colección presencial

7.1.1 Análisis estadísticos

En primera instancia se realiza un análisis estadístico de los datos, para ir conociendo tendencias y relaciones.

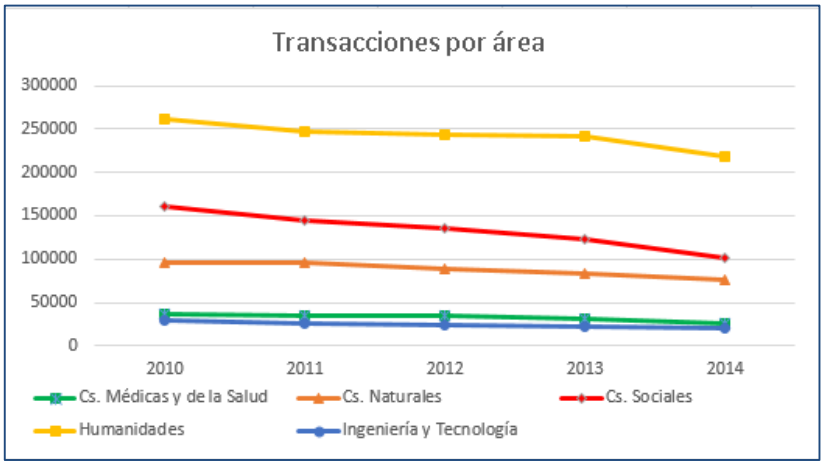

Imagen 7—1: Transacciones por área

Se observa que los materiales del área de Humanidades son los que registran la mayor cantidad de transacciones de los usuarios, seguida por Cs. Sociales, manteniéndose la tendencia durante los 5 años, con un leve descenso en el 2014.

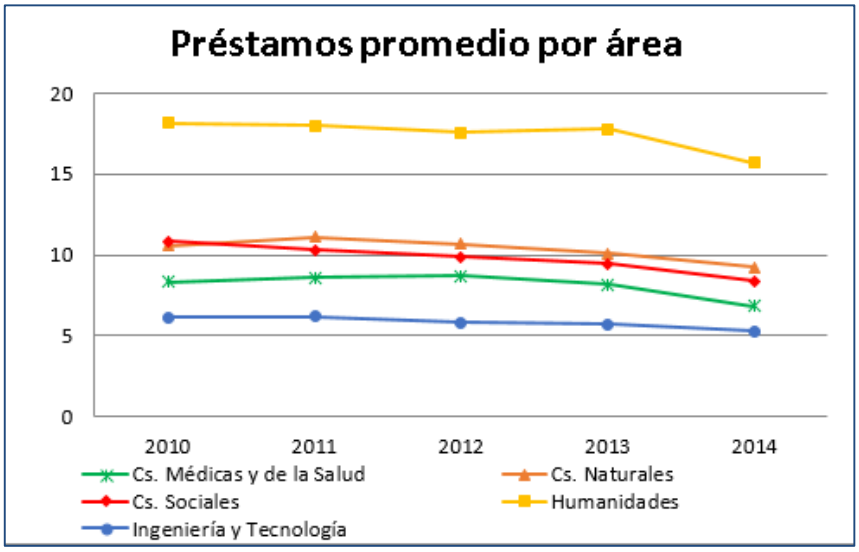

Imagen 7—2: Préstamos promedio por área

Al calcular los promedios por área, para descartar la influencia de aquellas áreas que fueron pedidas por una mayor cantidad de alumnos, se observa una tendencia similar, siendo el área de Humanidades la que posee el mayor promedio de transacciones, seguida por el área de Cs. Naturales y luego por Cs. Sociales.

- Los usuarios que piden material de Humanidades, Cs, naturales y Cs. Sociales piden en promedio más material que los que piden material de Cs. Médicas e Ingeniería.
- En todas las áreas se observa una disminución en la cantidad de material solicitado el último año, tendencia que es más leve en el área de Ingeniería.

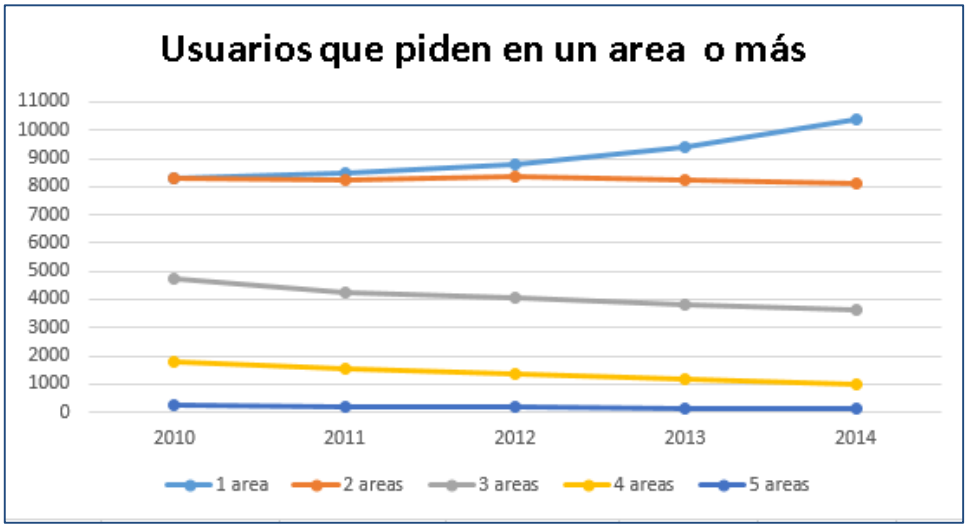

Imagen 7—3: Usuarios que piden en un área o más

Respecto a la multidisciplinariedad, se observa una fuerte tendencia en alza para solicitar material de un área en exclusiva, seguida por la tendencia de pedir en dos áreas, la que se mantiene muy similar en los cinco años.

La tendencia a pedir material de tres y cuatro áreas distintas presenta un leve descenso en los cinco años, se mantiene casi a la mitad de las solicitudes de una y dos áreas.

Los usuarios que piden material de las cinco áreas son muy pocos y la tendencia tiende a permanecer en los años. Se observa una correlación negativa, a mayor cantidad de áreas consultadas existe una menor cantidad de solicitudes.

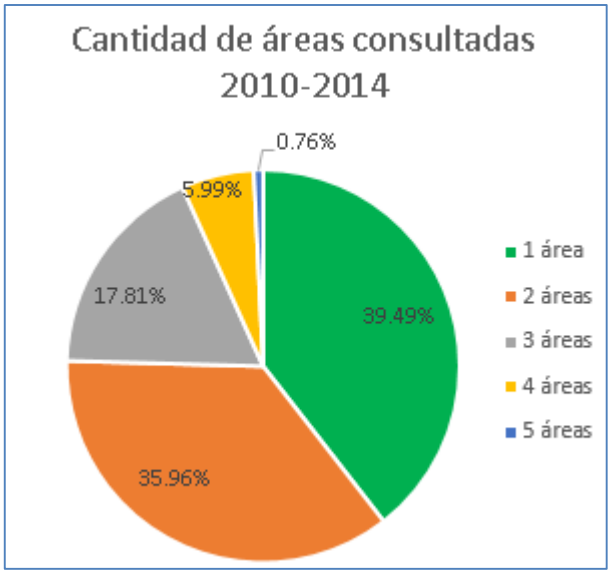

Imagen 7—4: Cantidad de áreas consultadas durante el 2010 al 2014, en %.

Si se compara en términos de porcentajes se observa que las solicitudes sólo de un área corresponden a casi un 40% del total de las consultas. Las solicitudes (2 y más áreas) consideradas en su conjunto corresponden a un 60% del total. Por lo tanto se podría hablar de una mayor frecuencia de consultas multidisciplinarias, si consideramos estas como la suma de consultas de 2 o más áreas.

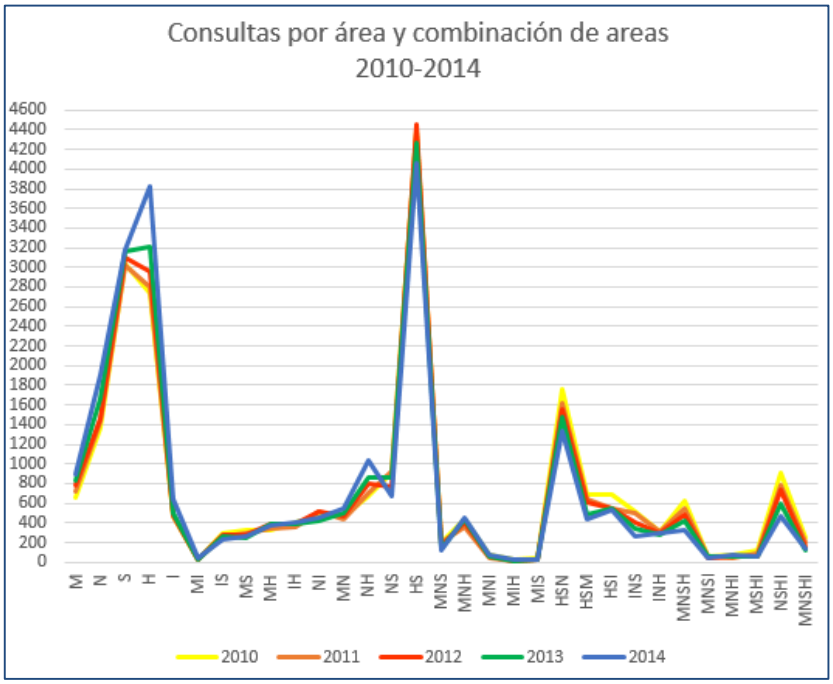

Imagen 7—5: Consultas por área y combinación de áreas

En el gráfico anterior se puede observar en detalle las consultas por áreas y cantidad de áreas a través de los cinco años de estudio. Se aprecia que Cs. Sociales (S) y Humanidades (H), tienen la mayor cantidad de consultas a través de los años y de las áreas exclusivas y combinadas. En general el año 2014 (en azul) tiene menos consultas que el resto de los años, con excepción de las áreas antes mencionadas, en donde se observa un aumento de las consultas de las áreas individuales. En donde también se observa un aumento es en la combinación MH, de Cs. Médicas y Humanidades, en NH de Cs. Naturales y Humanidades, y en MNH de Cs. Médicas, Naturales y Humanidades. Las combinaciones de consultas en 3 y 4 áreas, en general presentan una fuerte disminución de las consultas.

Respecto al tipo de usuario, los alumnos regulares son quienes realizan más consultas, principalmente del área de Humanidades, Cs. Sociales y Cs. Naturales, observándose una tendencia levemente decreciente. En las áreas menos solicitadas se observa una mayor estabilidad en la cantidad de consulta durante los 5 años.

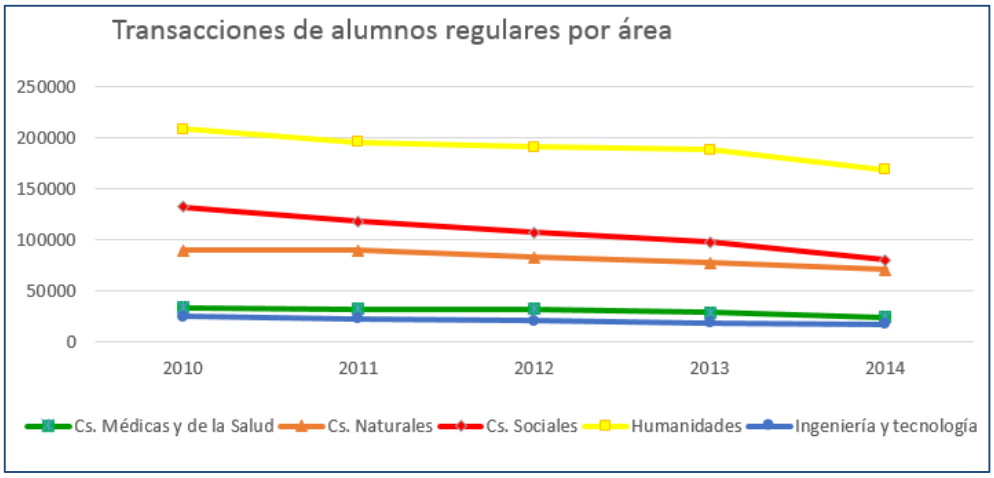

Imagen 7—6: Transacciones de alumnos regulares por área

Los docentes, y otros tipos de usuarios, presentan una tendencia similar en sus consultas, siendo los alumnos egresados quienes marcan una diferencia y prefieren el área de Cs. Sociales por sobre las otras.

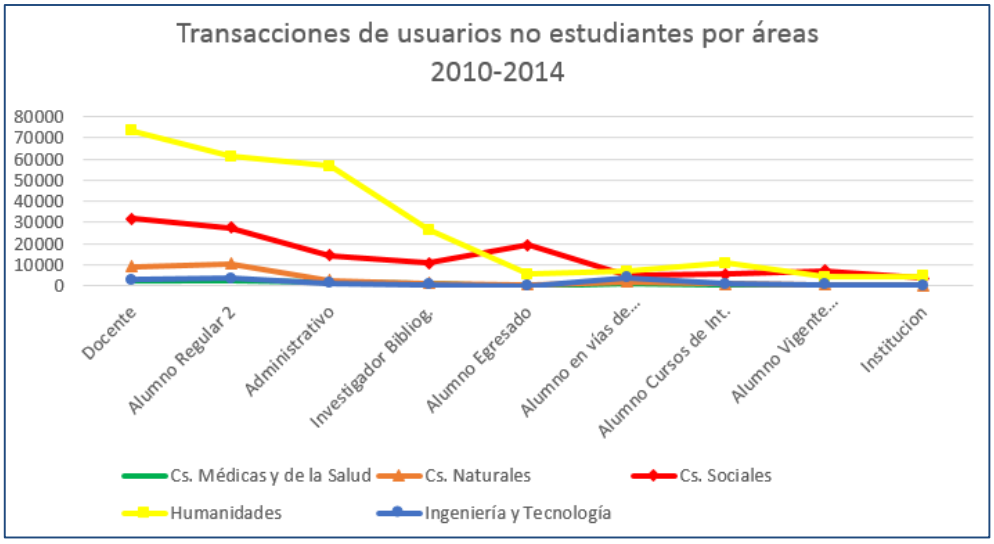

Imagen 7—7: Transacciones de usuarios no estudiantes por área

7.1.2 Conjuntos frecuentes de áreas consultadas

Para explorar si existe una relación entre las áreas que solicitan los usuarios, los datos se procesaron para establecer combinaciones de áreas frecuentes y reglas de asociación entre estas combinaciones.

Se realiza la generación de conjuntos de áreas que aparecen juntas con mayor frecuencia durante los cinco años de estudio, se utiliza el software Rapid Miner. Se trabaja con el modelo de asociación y conjuntos frecuentes, específicamente el operador FP-growth. Considerando un soporte mayor a 0.025. Se obtiene listas de conjuntos frecuentes de dominios para cada año, se puede observar un extracto de los conjuntos obtenidos en la siguiente tabla.

Tabla 7-1: Itemsets frecuentes 2010-2014

Soporte	item1	item2	item3	item4
0,607	Humanidades			
0,591	Cs. Sociales			
0,372	Cs. Naturales			
0,323	Humanidades	Sociales		
0,196	Humanidades	Cs. Naturales		
0,190	Cs. Sociales	Cs. Naturales		
0,180	Ingenieria y Tecnologia			
0,177	Cs. Medicas y de la Salud			
0,112	Humanidades	Cs. Sociales	Cs. Naturales	
0,100	Humanidades	Ingeniería y Tecnología		
0,097	Cs. Naturales	Ingenieria y Tecnologia		
0,095	Humanidades	Cs. Medicas y de la Salud		
0,083	Cs. Naturales	Cs. Medicas y de la Salud		
0,080	Cs. Sociales	Cs. Medicas y de la Salud		
0,066	Humanidades	Cs. Sociales	Ingenieria y Tecnologia	
0,065	Cs. Sociales	Ingenieria y Tecnologia		
0,058	Cs. Sociales	Cs. Naturales	Ingenieria y Tecnologia	
0,054	Humanidades	Cs. Naturales	Ingenieria y Tecnologia	
0,050	Humanidades	Cs. Naturales	Cs. Medicas y de la Salud	
0,038	Cs. Sociales	Cs. Naturales	Cs. Medicas y de la Salud	
0,038	Humanidades	Cs. Sociales	Cs. Naturales	Ingenieria y Tecnologia
0,025	Humanidades	Cs. Sociales	Cs. Naturales	Cs. Medicas y de la Salud
0,006	Ingenieria y Tecnologia	Cs. Medicas y de la Salud		

Las áreas de Humanidades, Cs. Sociales y Cs. Naturales presentan la mayor frecuencia en las consultas, tanto solas como combinadas entre ellas. La combinación de la dupla Humanidades-Cs. Sociales es la más frecuente, seguida por Humanidades-Cs. Naturales y Cs. Sociales-Cs. Naturales. El trío Humanidades, Cs. Sociales y Cs. Naturales es el más frecuente, seguido tras varios puntos de diferencia por Humanidades, Cs. Sociales e Ingeniería.

Por otro lado, las consultas menos frecuente son las de la combinación Ingeniería y Cs. Médicas, que es la más esporádica, con un soporte de 0,006. Seguida por los cuartetos Humanidades, Cs. Sociales, Cs. Naturales y Cs. Médicas, y Humanidades, Cs. Sociales, Cs. Naturales e Ingeniería.

Se observa que el aumento de la frecuencia va ligado a la aparición de alguna de las tres áreas Humanidades, Cs. Sociales y Cs. Naturales, y que la disminución de la frecuencia va con las áreas Ingeniería y Cs. Médicas.

Durante los años 2010-2011 es el área de Cs. Sociales la que presenta la mayor frecuencia seguida por Humanidades, y los años 2012 a 2014 esta frecuencia se invierte. Lo que lleva a preguntarnos ¿Existe un patrón en el tiempo respecto a las consultas de determinadas áreas?

7.1.3 Entropía de las consultas por área

Para determinar la homogeneidad / heterogeneidad de las consultas en las distintas áreas durante los cinco años de estudios, se calculó la entropía de las consultas para cada área y combinación de ellas durante el tiempo de estudio, cuyos resultados se observan en los siguientes gráficos. En donde M=Medicina, S=Cs. Sociales, N=Cs. Naturales, I=Ingeniería y H=Humanidades.

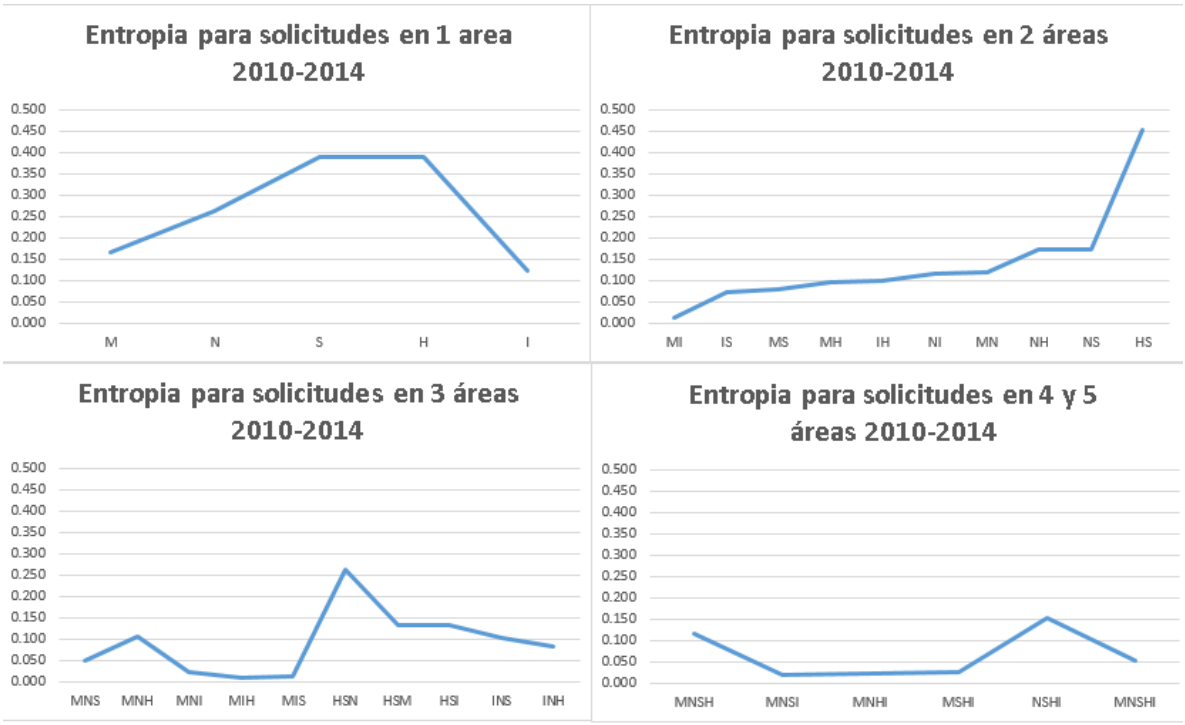

Imagen 7—8: Entropía por áreas

Para el cálculo de entropía se consideró la probabilidad de ocurrencia de cada área y sus combinaciones, respecto al total de las consultas para cada año. Luego se aplicó la siguiente formula, y se sumó los resultados de cada año por área.

$$H = -p_1 \log_2(p_1) - p_2 \log_2(p_2) - \ldots - p_k \log_2(p_k) = -\sum_{i=1}^{k} p_i \log_2(p_i)$$

Se observa que Cs. Sociales y Humanidades mantienen una alta entropía en las consultas exclusivas, mientras que Humanidades tiene una baja entropía, lo que se puede explicar por las variaciones explicadas anteriormente, pues ha ido aumentando fuertemente la cantidad de consultas exclusivas de Humanidades (de 2700 el 2010 a más de 3800 el 2014), y Cs. Sociales se ha mantenido siempre en un rango aproximado de 3.100 consultas anuales.

Naturales también tiene una baja entropía al ser consultada en exclusiva, ya que en los 5 años ha presentado un aumento permanente de aproximadamente un 30%.

Las áreas de Ingeniería y Medicina presentan una baja entropía, pues presenta una variación de las consultas, bajas y altas, durante los años a pesar de mantenerse siempre bajo las 1000.

En las consultas de áreas combinadas observamos que aquellas que incluyen la dupla HS (Sociales-Humanidades), mantienen una alta entropía, lo que nos indica que ambas se mantienen con una alta cantidad de consultas a través de los años, y si se asocian a áreas con baja entropía en forma exclusiva, suben la entropía del conjunto. Por otro lado la dupla MI (Medicina-Ingeniería) tiende a estar presente en aquellas combinaciones con baja entropía, pues presentan variaciones en la cantidad de consultas en el tiempo.

Al observar las relaciones entre estas áreas surge la inquietud de identificar algún patrón de asociación entre ellas.

7.1.4 Reglas de asociación por áreas

Se generaron reglas de asociación utilizando el software Rapid Miner, con el modelo de asociación y conjuntos frecuentes, específicamente los operadores FP-growth y Create Association Rules. Se consideró un soporte mínimo de 0.15.

Se continúa el análisis de los dominios aplicando técnicas de minería de datos, se utiliza el software Rapid Miner para determinar reglas de asociación de dominios consultados en los

tres años. Se utiliza el modelo de asociación y conjuntos frecuentes, específicamente los operadores FP-growth y Create Association Rules.

Al generar las reglas de asociación con una confianza mínima de 0.1, y un soporte de 0.025, se obtienen reglas similares para los cinco años, en el cuadro a continuación se muestran las 10 primeras para cada año.

Tabla 7-2: Reglas de asociación por áreas

Premisa	Conclusión	2010		2011		2012		2013		2014	
		Soporte	Conf.	Soporte	Conf.	Soporte	Conf.	Soporte	Conf.	Soporte	Conf.
Cs. Sociales	Humanidades	0,40	0,64	0,39	0,63	0,38	0,63	0,35	0,61	0,32	0,61
Humanidades	Cs. Sociales	0,40	0,65	0,39	0,64	0,38	0,62	0,35	0,59	0,32	0,53
Cs. Sociales	Cs. Naturales	0,22	0,35	0,21	0,34	0,19	0,32	0,18	0,31	0,15	0,28
Cs. Naturales	Cs. Sociales	0,22	0,58	0,21	0,56	0,19	0,52	0,18	0,48	0,15	0,41
Humanidades	Cs. Naturales	0,22	0,35	0,20	0,33	0,20	0,32	0,19	0,31	0,18	0,30
Cs. Naturales	Humanidades	0,22	0,55	0,20	0,53	0,20	0,53	0,19	0,51	0,18	0,50
Cs. Sociales	Humanidades, Cs. Naturales	0,15	0,24	0,14	0,22	0,13	0,21	0,12	0,20	0,10	0,19
Humanidades	Cs. Sociales, Cs. Naturales	0,15	0,25	0,14	0,23	0,13	0,21	0,12	0,19	0,10	0,16
Cs. Sociales, Humanidades	Cs. Naturales	0,15	0,38	0,14	0,36	0,13	0,34	0,12	0,33	0,10	0,31
Cs. Naturales	Cs. Sociales, Humanidades	0,15	0,39	0,14	0,36	0,13	0,35	0,12	0,32	0,10	0,31

Para todos los años las reglas con más alto soporte y confianza son aquellas que asocian en premisas y conclusiones las consultas en áreas de Cs. Sociales, Humanidades y Cs. Naturales.

Si consideramos la certeza (confianza) de que dos áreas se consulten juntas, las reglas de asociación obtenidas nos indican que para el 2010 y 2011 lo más probable es que cuando un usuario solicitó material de Humanidades pidió además material de Cs. Sociales y en segundo lugar, con una probabilidad mucho más baja, de Cs. Naturales.

Para el 2012-2014 la confianza más alta se presenta con la premisa de que cuando un usuario pidió material de Cs. Sociales, la conclusión más probable es que pidió además algo de Humanidades, y más baja está la probabilidad de pedir algo de Cs. Naturales.

Si un usuario pidió algo de Cs. Naturales, durante los años 2010-2011 solicitó algo de Cs. Sociales en primera opción, y en segunda opción algo de Humanidades. Es interesante notar que para estas últimas reglas existe una confianza similar, pero que va a revertirse a partir del 2012, cuando quienes pidieron material de Cs. Naturales pidieron con mayor confianza material de Humanidades y luego de Cs. Sociales, haciendo que la diferencia entre el indicador de confianza entre ambas reglas aumente de 1 punto a 9.

Respecto a la certeza de que tres áreas se consulten juntas, la regla con más alta confianza, a lo largo de los 5 años, es la que nos presenta la premisa que cuando se pidió material de Cs. Naturales, se concluye que se pidió también material de Cs. Sociales y Humanidades. Seguida por la premisa que quienes pidieron material de Cs. Sociales y Humanidades, se concluye que solicitaron material de Cs. Naturales.

En definitiva son estas tres áreas, Humanidades, Cs. Sociales y Cs. Naturales, las que representan una alta asociación multidisciplinaria.

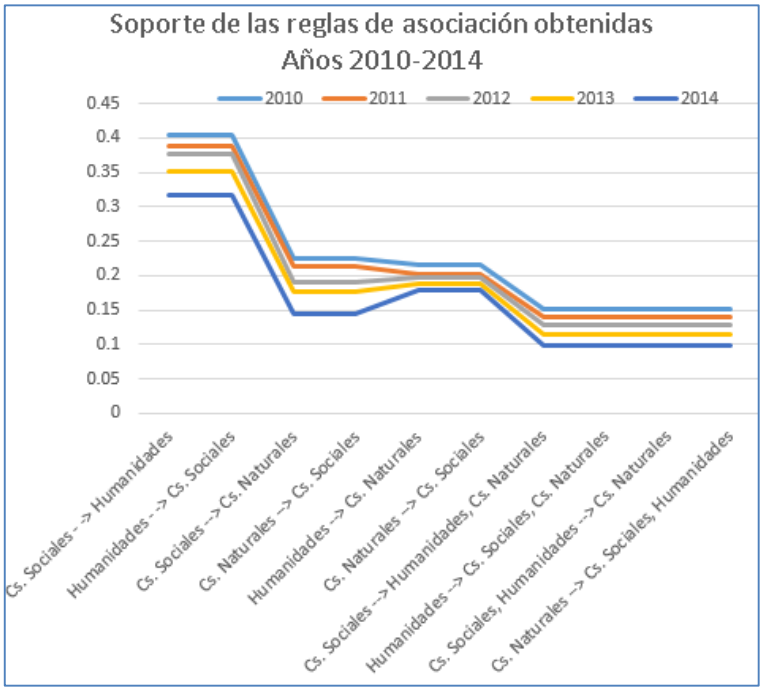

Imagen 7—9: Soporte de reglas de asociación

Todas las reglas encontradas presentan una baja sostenible del soporte y la confianza en los 5 años estudiados, la que es más acentuada en la regla de Cs. Sociales - Humanidades, y menor en las reglas de Humanidades - Cs. Naturales y Cs. Naturales – Humanidades.

Es interesante detenerse en el comportamiento de las consultas relacionadas con el área de Cs. Naturales, ya que muestra un aumento de las consultas exclusivas en el tiempo de estudio y un aumento de las consultas asociadas con Cs. Médicas y Humanidades.

Lo anterior, junto al aumento de las consultas de un área exclusiva y la disminución de las consultas de varias disciplinas, podría interpretarse como una tendencia a las consultas de sólo una disciplina, a lo más dos, de la colección.

7.1.5 Consultas por área del material y área del usuario

Al comparar las consultas por áreas del usuario y áreas del material se observa una tendencia a la consulta mayoritaria de material de su área en usuarios del área de Cs. Naturales, Cs. Médicas y de la Salud y Humanidades. Los usuarios de Cs. Sociales consultan en primer lugar material de Humanidades y en segundo lugar de su área. Los de Ciencia y Tecnología consultan material de Cs. Naturales y Humanidades más que de su propia área.

Tabla 7-3: Promedio de consultas por área del material y área del usuario 2012-2014

Area del usuario	Transacciones por Area del material 2012-2014					Total
	Cs. Medicas y de	Cs. Naturales	Cs. Sociales	Humanidades	Ingeniería y Tecnologia	
Ciencia Naturales	14,90	52,90	8,85	24,10	4,47	21,04
Ciencia y Tecnologia	4,11	21,78	10,55	15,22	12,07	12,75
Ciencias Medicas y de la S	33,93	13,61	5,25	20,77	3,09	15,33
Ciencias Sociales	6,52	13,24	24,71	32,45	5,37	16,46
Humanidades	3,69	5,19	22,64	104,65	5,37	28,31
Total general	12,63	21,35	14,40	39,44	6,07	18,78

Los usuarios que consultan material por sobre el promedio de la Universidad, 18,78, son los del área de Humanidades, con 28, seguidos por los de Cs. Naturales, con 21. El resto de las áreas se mantienen por debajo del promedio.

Los materiales más consultados pertenecen al área de Humanidades, con un promedio de consultas de 39,44, y a la de Cs. Naturales con un promedio de 21,35. El resto de las áreas se mantienen por debajo del promedio.

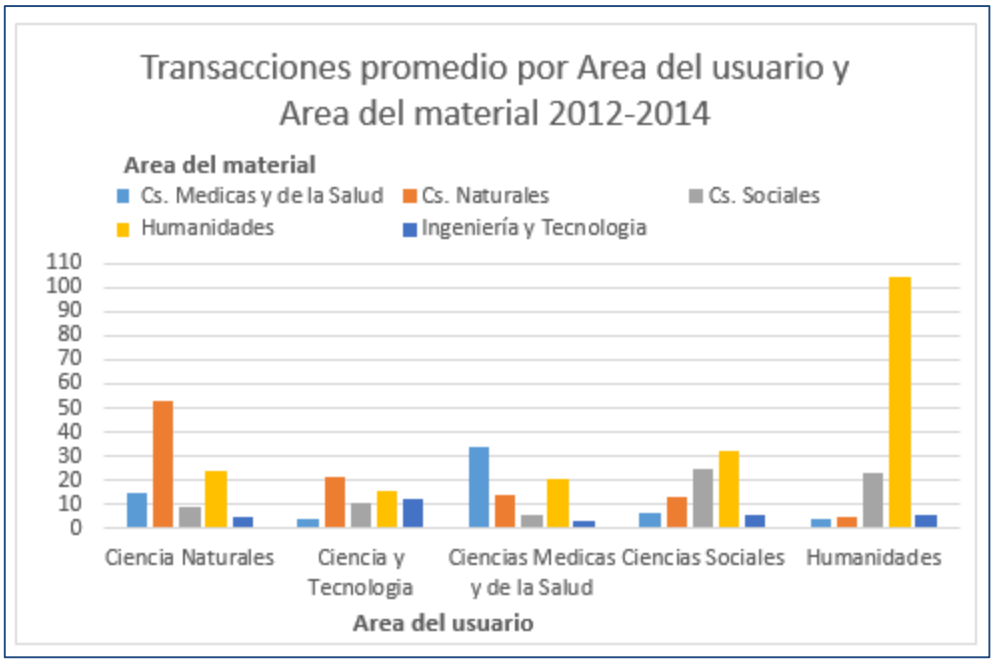

Imagen 7—10: Transacciones promedio por área del usuario y área del material

Respecto a las consultas de material de distintas áreas, por un mismo usuario, se puede observar que la mayoría de los usuarios consultan material de 2 o 1 área, con 38,9% y 35,4% respectivamente, las consultas de 3 áreas distintas bajan a un 17,5%, y las de 4 áreas corresponden a un 5,9% y las de 5 áreas sólo son un 0.75%.

Tabla 7-4: Cantidad de áreas consultadas por área del usuario

	Cantidad de áreas consultadas					
Area del usuario	1	2	3	4	5	Total general
Ciencia Naturales	1048	1296	885	338	41	3608
Ciencia y Tecnologia	4104	3650	2430	1117	156	11457
Ciencias Medicas y de la S	1585	1554	906	299	36	4380
Ciencias Sociales	8244	8059	3662	890	74	20929
Humanidades	2127	2596	895	185	20	5823
Total general	17108	17155	8778	2829	327	46197

Los usuarios de Ciencia y Tecnología son quienes presentan una mayor cantidad de consultas de material de 5 y 4 áreas distintas. Los usuarios de Ciencias Sociales y de Ciencia y Tecnología son quienes presentan mayor cantidad de consultas de material de tres áreas distintas. De Ciencias Sociales son los usuarios que piden mayoritariamente material de 1 o 2 áreas.

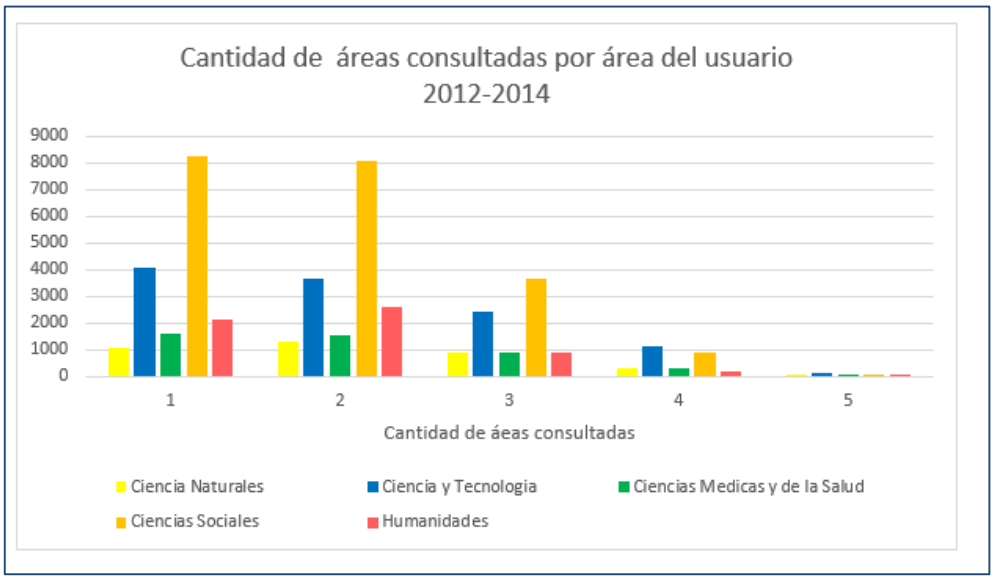

Imagen 7—11: Cantidad de áreas consultadas por área del usuario

En el gráfico anterior se observa que prevalecen las consultas a 1 y 2 áreas, especialmente en usuarios de Ciencias Sociales y de Ciencia y Tecnología.

En Ciencias Naturales y Humanidades se observa un leve aumento en la cantidad de consultas en dos áreas respecto a una sola. Cs. Médicas se mantiene con consultas en 1 o 2 áreas.

Las consultas de 3 áreas bajan notablemente para todos los usuarios, esta baja se hace más notoria en las consultas de 4 y 5 áreas.

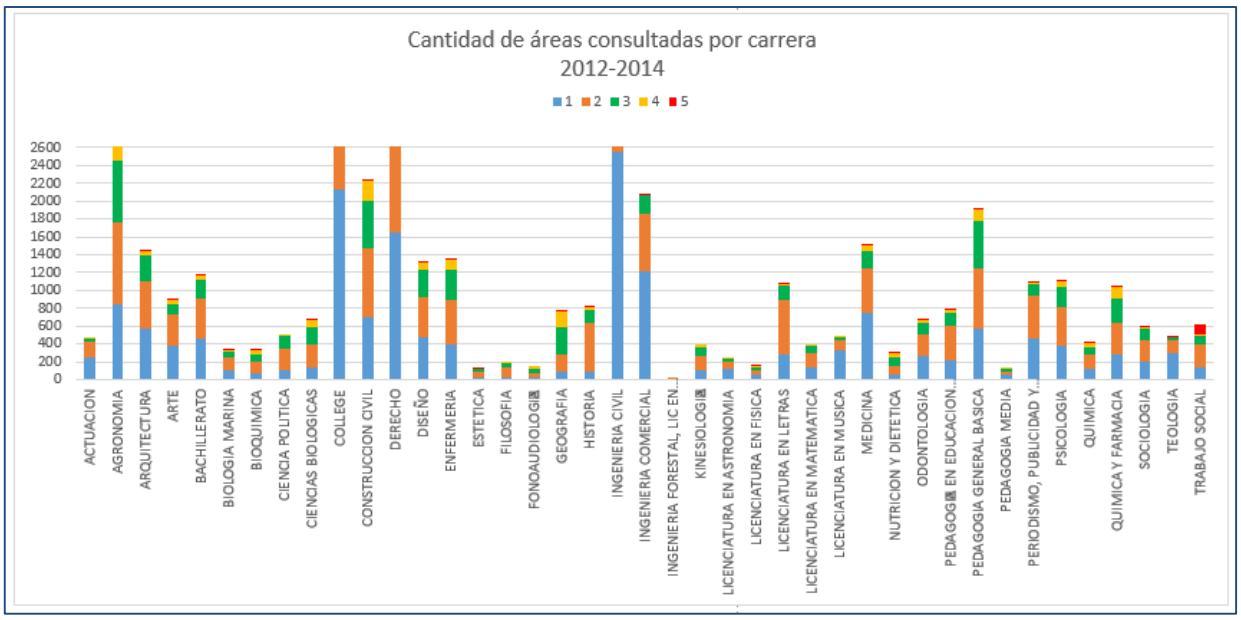

Imagen 7—12: Cantidad de áreas consultadas por carrera

Al realizar el análisis por carrera se observa que entre las carreras que más transacciones realizan, las carreras que consultan principalmente un área son College e Ingeniería Civil. Derecho consulta de 2 áreas, con 4 y 5 áreas se destacan Agronomía, Arquitectura, Bachillerato, Construcción civil, Diseño, Enfermería, Ing. Comercial, Medicina, Pedagogía Básica, Periodismo, Psicología y Química y Farmacia.

Se observa que la mayoría de las carreras presentan consultas de distintas disciplinas. No existen carreras que pidan exclusivamente de un área.

Para comparar las frecuencias observadas en las consulta de material de áreas distintas de la que cubre la Unidad Académica a la que pertenece el usuario, se realiza un test binomial, considerando los años 2010 al 2014, cuyos resultados se incluyen en la siguiente tabla.

Se considera las hipótesis:

H_0: El 50% de las consultas son multidisciplinarias

H_1: Menos del 50% de las consultas son multidisciplinarias

Tabla 7-5: Test binomial de consultas en áreas distintas a la de la Unidad Académica

Test de proporciones		2010			2011			2012			2013			2014		
Unidad Académica	Area	P de éxito	p-value	T*	P de éxito	p-value	T*	P de éxito	p-value	T*	P de éxito	p-value	T*	P de éxito	p-value	T*
Esc. de Const. Civil	C. y Tecnologia	79,97%	2.2e-16	>	84,11%	2.2e-16	>	82,98%	2.2e-16	>	87,57%	2.2e-16	>	81,48%	2.2e-16	>
Esc. de Ingenieria	C. y Tecnologia	76,81%	2.2e-16	>	77,05%	2.2e-16	>	78,03%	2.2e-16	>	77,47%	2.2e-16	>	75,71%	2.2e-16	>
Esc. de Agronomia	C. y Tecnologia	74,97%	2.2e-16	>	75,21%	2.2e-16	>	78,90%	2.2e-16	>	80,70%	2.2e-16	>	81,24%	2.2e-16	>
Esc. de Enfermeria	Cs. Médicas y d	42,02%	2.2e-16	<	45,52%	2.2e-16	<	46,74%	3.08e-13	<	45,73%	2.2e-16	<	51,52%	0.006514	>
Esc. de Medicina	Cs. Médicas y d	39,18%	2.2e-16	<	34,90%	2.2e-16	<	35,80%	2.2e-16	<	41,41%	2.2e-16	<	45,44%	2.2e-16	<
Fac. de Fisica	Cs. Naturales	22,67%	2.2e-16	<	23,75%	2.2e-16	<	26,19%	2.2e-16	<	24,28%	2.2e-16	<	27,46%	2.2e-16	<
Fac. de Matematica	Cs. Naturales	21,36%	2.2e-16	<	18,13%	2.2e-16	<	20,54%	2.2e-16	<	23,47%	2.2e-16	<	25,53%	2.2e-16	<
Fac. de Quimica	Cs. Naturales	45,30%	2.2e-16	<	41,17%	2.2e-16	<	41,39%	2.2e-16	<	39,61%	2.2e-16	<	44,08%	2.2e-16	<
Fac. de Cs. Biologica	Cs. Naturales	47,01%	2.2e-16	<	49,66%	0.3521	<	45,85%	2.2e-16	<	43,99%	2.2e-16	<	47,65%	5,15E-06	<
Fac. de Cs. Economic	Cs. Sociales	34,72%	2.2e-16	<	36,57%	2.2e-16	<	37,40%	2.2e-16	<	39,97%	2.2e-16	<	39,16%	2.2e-16	<
Bachillerato	Cs. Sociales	67,18%	2.2e-16	>	78,36%	2.2e-16	>	85,00%	4,83E-10	>	53,66%	0.5811	>	44,90%	0.5682	<
Colleges	Cs. Sociales	74,04%	2.2e-16	>	71,80%	2.2e-16	>	72,46%	2.2e-16	>	74,59%	2.2e-16	>	79,06%	2.2e-16	>
Fac. de Derecho	Cs. Sociales	25,38%	2.2e-16	<	26,40%	2.2e-16	<	26,98%	2.2e-16	<	26,65%	2.2e-16	<	29,24%	2.2e-16	<
Fac. de Educacion	Cs. Sociales	46,37%	2.2e-16	<	46,60%	2.2e-16	<	49,79%	0.4328	<	57,52%	2.2e-16	>	61,80%	2.2e-16	>
Fac. de Comunicacio	Cs. Sociales	68,12%	2.2e-16	>	69,75%	2.2e-16	>	69,82%	2.2e-16	>	73,02%	2.2e-16	>	71,54%	2.2e-16	>
Esc. de Psicologia	Cs. Sociales	51,87%	1.14e-07	>	51,24%	0.001869	>	56,54%	2.2e-16	>	56,60%	2.2e-16	>	55,81%	2.2e-16	>
Esc. de Trabajo Soci	Cs. Sociales	32,24%	2.2e-16	<	36,45%	2.2e-16	<	91,22%	2.2e-16	>	43,69%	2.2e-16	<	37,38%	2.2e-16	<
Inst. de C. Politica	Cs. Sociales	38,95%	2.2e-16	<	37,44%	2.2e-16	<	87,05%	2.2e-16	>	43,34%	2.2e-16	<	41,66%	2.2e-16	<
Inst. de Geografia	Cs. Sociales	69,81%	2.2e-16	>	73,24%	2.2e-16	>	45,29%	2,78E-02	<	68,09%	2.2e-16	>	69,63%	2.2e-16	>
Inst. de Sociologia	Cs. Sociales	51,58%	0.000289	>	51,20%	0.01176	>	55,47%	2.2e-16	>	55,32%	2.2e-16	>	52,45%	2.03e-05	>
Esc. de Arquitectura	Cs. Sociales	83,94%	2.2e-16	>	87,38%	2.2e-16	>	86,53%	2.2e-16	>	87,02%	2.2e-16	>	86,51%	2.2e-16	>
Inst. de Est. Urbanos	Cs. Sociales	33,61%	2.2e-16	<	9,57%	2.2e-16	<	45,29%	2,78E-02	<	44,88%	2,86E-04	<	39,71%	2.2e-16	<
Esc. de Arte	Humanidades	8,95%	2.2e-16	<	7,68%	2.2e-16	<	11,43%	2.2e-16	<	11,33%	2.2e-16	<	10,81%	2.2e-16	<
Esc. de Teatro	Humanidades	7,86%	2.2e-16	<	7,68%	2.2e-16	<	8,86%	2.2e-16	<	9,76%	2.2e-16	<	7,18%	2.2e-16	<
Fac. de Teologia	Humanidades	5,19%	2.2e-16	<	6,07%	2.2e-16	<	5,22%	2.2e-16	<	5,70%	2.2e-16	<	3,97%	2.2e-16	<
Inst. de Estetica	Humanidades	15,65%	2.2e-16	<	15,22%	2.2e-16	<	12,91%	2.2e-16	<	16,85%	2.2e-16	<	15,42%	2.2e-16	<
Inst. de Historia	Humanidades	3,84%	2.2e-16	<	36,64%	2.2e-16	<	37,36%	2.2e-16	<	38,39%	2.2e-16	<	40,05%	2.2e-16	<
Esc. de Diseño	Humanidades	29,63%	2.2e-16	<	30,04%	2.2e-16	<	34,67%	2.2e-16	<	35,61%	2.2e-16	<	36,78%	2.2e-16	<
Inst. de Letras	Humanidades	13,25%	2.2e-16	<	12,72%	2.2e-16	<	13,70%	2.2e-16	<	13,98%	2.2e-16	<	15,26%	2.2e-16	<
Inst. de Filosofia	Humanidades	17,57%	2.2e-16	<	20,12%	2.2e-16	<	27,61%	2.2e-16	<	19,79%	2.2e-16	<	15,54%	2.2e-16	<
Inst. de Musica	Humanidades	4,66%	2.2e-16	<	4,05%	2.2e-16	<	4,72%	2.2e-16	<	4,37%	2.2e-16	<	3,94%	2.2e-16	<

T*= Tendencia respecto a la H1

Se observa en color amarillo el porcentaje de los usuarios que cumplen con la hipótesis nula y presentan un porcentaje por sobre el 50% de consultas de áreas distintas a las de su Unidad Académica. Estos corresponden a Unidades del área de Ciencia y Tecnología, Bachillerato, College, Geografía, Sociología y Arquitectura. Esta mayor multidisciplinariedad de sus consultas se podría explicar si se considera que son carreras con una base de formación común que incluye ramos de áreas distintas a las de la Unidad Académica (por ej. Cálculo, Química, Física). En color naranja se representa a los usuarios que cumplen con la hipótesis alternativa, pues tienen un porcentaje menor al 50% de consultas de áreas distintas a las de su Unidad Académica. Corresponden a la mayoría de los usuarios, de áreas como Ciencias Médicas y de la Salud, Derecho, Educación, Trabajo Social, Ciencia Política, Estudios Urbanos y todas las carreras del área de Humanidades.

Existen algunas pocas unidades que modifican su tendencia, por ejemplo Enfermería y Educación han aumentado la multidisciplinariedad de sus consultas en los últimos años de estudio. Trabajo Social y Ciencia Política aumentaron por el año 2012 las consultas multidisciplinarias pero luego volvieron a bajar.

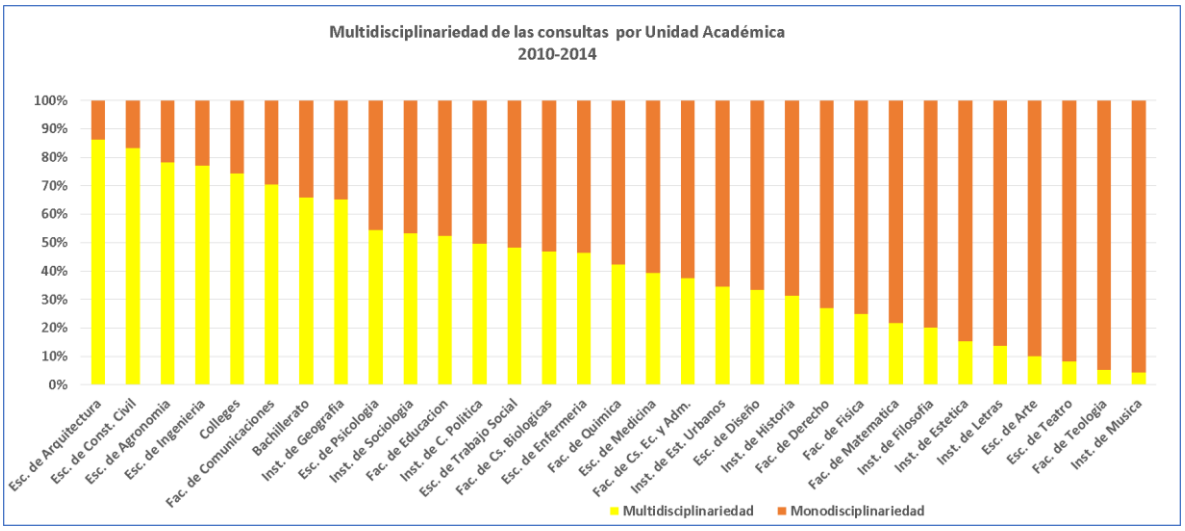

Imagen 7—13: Multidisciplinariedad de las consultas por Unidad Académica

En el gráfico anterior se observa que las unidades académicas con mayor cantidad de consultas multidisciplinarias son Arquitectura, Construcción Civil, Agronomía e Ingeniería. Las unidades con consultas mayoritariamente monodisciplinarias corresponden a Música, Teología, Teatro, Arte y Letras. Las unidades con consultas alrededor del 50% multidisciplinarias son Trabajo Social, Cs. Biológicas, Enfermería y Química.

Durante los años 2010 y 2014, entre un 68% y un 61% de las unidades académicas tenía una mayoría de consultas monodisciplinarias, en promedio consultaban un 57% de material de sus áreas y un 43% de otras áreas.

Tabla 7-6: Proporción de multidisciplinariedad de las consultas

	2010	2011	2012	2013	2014
>50% multidisciplinariedad	32,26%	32,26%	35,48%	35,48%	38,71%
<50% multidisciplinariedad	67,74%	67,74%	64,52%	64,52%	61,29%

En general se observa en los cinco años de estudio que existe una tendencia a tener consultas multidisciplinarias bajo el 50%, en al menos el 61% de las Unidades Académicas de la Universidad. Como máximo el 39% de las Unidades Académicas presenta consultas multidisciplinarias por sobre el 50%.

Por lo tanto, se rechaza la hipótesis nula ya que existe una diferencia estadística significativa con los resultados obtenidos, la mayoría de las consultas por Unidad Académica en áreas distintas a la de la unidad tienen un p-value <0.05. Más del 60% de las Unidades Académicas tienen consultas multidisciplinarias.

7.1.6 Uso de la biblioteca y notas

Para comenzar a trabajar con el set de datos generado en el Proyecto Fondecyt 1131109, se realiza un análisis estadístico de los datos, centrando la atención en la información del rendimiento académico de los estudiantes y en el área o carrera de la que provienen, en relación con la cantidad de transacciones que han tenido en los años 2012 y 2014.

En primer lugar se revisa la cantidad de transacciones que presentan los usuarios y su relación con las notas. Para representar las notas se utiliza el cálculo del Zscore de notas por carrera, para normalizar las notas en todas las carreras.

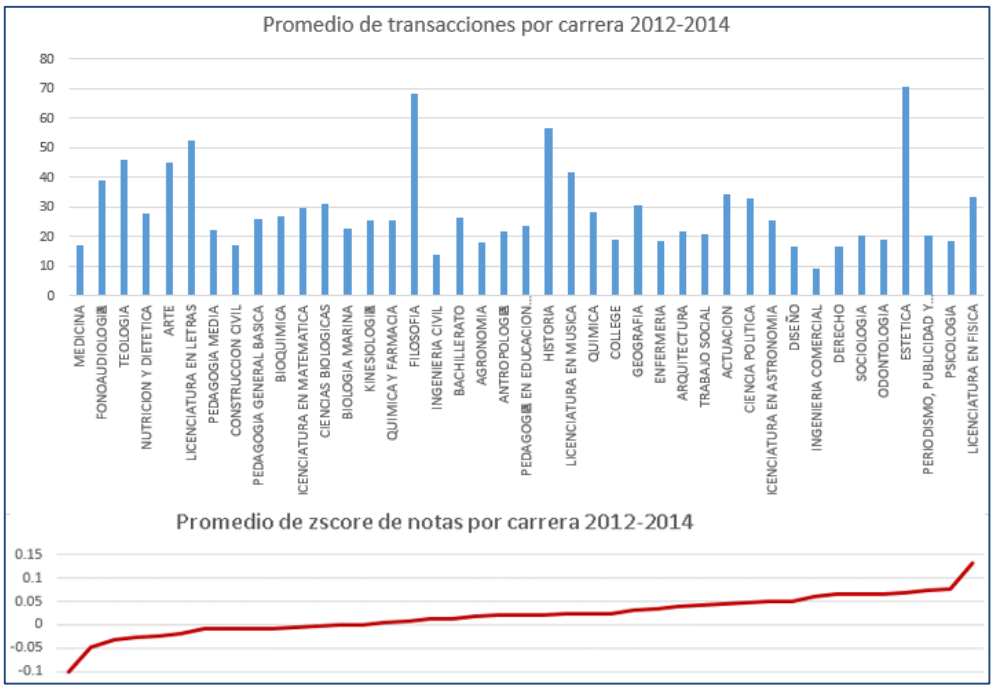

Imagen 7—14: Promedio de transacciones y Zscore por carrera 2012-2014

En el gráfico anterior se aprecia una leve relación entre cantidad de transacciones y notas de los alumnos, el promedio de transacciones por alumno es cercano a 13, las carreras con notas más altas en general superan el promedio (con la excepción de Ingeniería Comercial). También las carreras con menor Zscore superan el promedio de transacciones.

Las carreras que más consultan o piden material, por sobre 30 en promedio, son Estética, Filosofía, Historia, Letras, Teología, Arte, Música, y luego, Fonoaudiología, la primera del área de la Salud, Actuación, Física, la primera del área de las Ciencias Exactas, seguida por Ciencias Biológicas, Ciencias Políticas y Geografía.

Las carreras que menos transacciones han solicitado, menos de 13 en promedio, son Ingeniería Comercial e Ingeniería Civil. Las carreras que tienen menos de 20 transacciones en promedio son Diseño, Derecho, Medicina, Construcción Civil, o, Psicología, Enfermería, Odontología y College. Las carreras tienden a permanecer en el promedio del Zscore, cercano a cero.

A simple vista se observa una tendencia en que las careras humanistas son quienes tienen mayor cantidad de consultas de material de la Biblioteca, seguida por las carreras de Ciencias Exactas, y quienes menos piden material son las carreras de Ingeniería y Cs. De la Salud. A continuación verificaremos estas apreciaciones, agrupando las carreras por áreas OCDE.

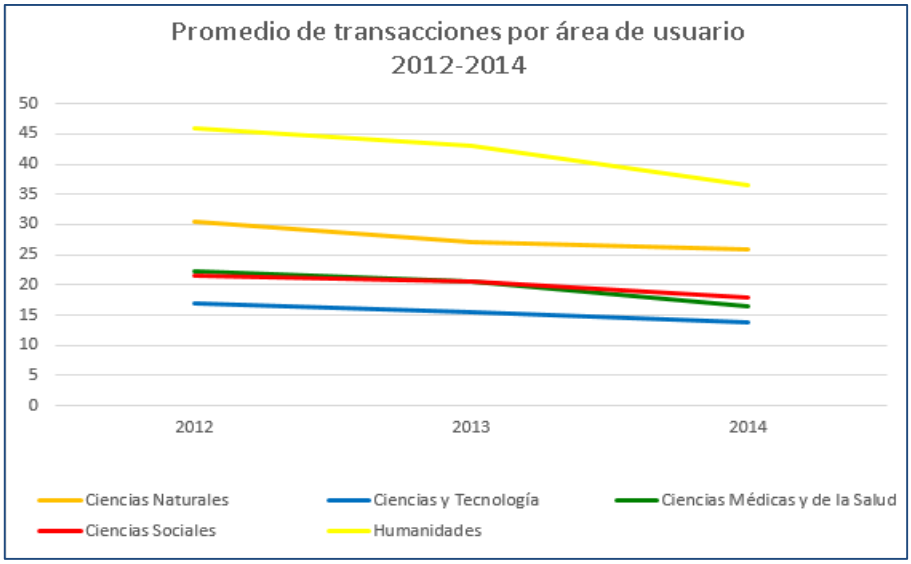

Imagen 7—15: Promedio de transacciones por área de usuario

Se observa efectivamente que las carreras del área de Humanidades son las que presentan la mayor cantidad de consultas, seguidas por las Cs. Naturales, luego vienen las de Ciencias Sociales y Cs. Médicas, para finalizar con las Ciencias y Tecnología.

En todas las áreas se aprecia una tendencia a la disminución de las consultas en el tiempo de estudio. Siendo más fuerte la baja en las áreas con mayor cantidad de transacciones y más leve en las áreas que piden menos.
A continuación revisaremos la dispersión de transacciones y notas de las distintas carreras.

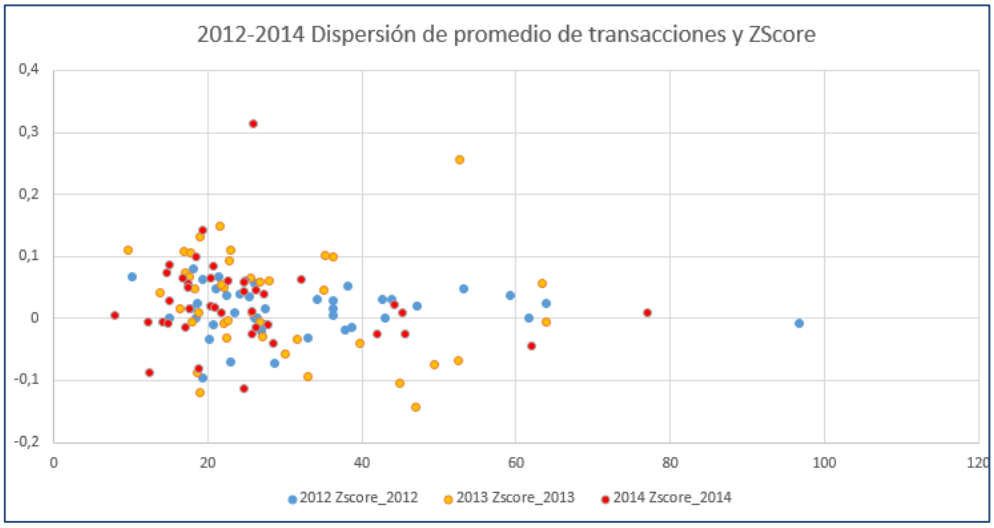

Imagen 7—16: Dispersión de promedio de transacciones y Zscore

En el gráfico anterior, donde cada punto representa una carrera y cada color un año de estudio, al comparar el promedio de transacciones con las notas se observa que existe una mayor densidad en aquellas carreras que piden entre 10 y 30 transacciones en promedio, y tienen un zscore de -0.12 y 0.15. Además se puede observar otra área con menor densidad entre aquellas carreras con un promedio de transacciones sobre 30 y un Zscore entre -0.1 y 0.1. Lo que nos indica que la mayoría de la población en estudio pide menos de 30 y tiene un Zscore entre -0.12 y 0.15, existiendo un grupo minoritario que pide sobre 30 transacciones y se mueve en un rango de Zscore un poco menor, pero similar.

Se observa una mayor densidad en las transacciones menores a 30, en los tres años de estudio. Luego se ve un rango de menor densidad entre las 30 a 60 consultas, especialmente en los años 2012 y 2013. Las transacciones han disminuido levemente

desde el 2012 al 2014, especialmente no se observan tantos casos fuera de rango y la población tiende a concentrarse en el rango de menor cantidad de transacciones.

Se realiza una prueba de proporciones para determinar la diferencia de proporciones de la cantidad de usuarios con notas sobre el promedio que piden material en biblioteca y de aquellos usuarios con notas bajo el promedio que piden material en biblioteca, los resultados se muestran en la siguiente tabla, que se obtuvo utilizando el programa R.

Las hipótesis contrastadas son:

$H_{o:}$ No existe diferencia entre el rendimiento académico de los alumnos que usan los servicios de la biblioteca por sobre el promedio.

$H_{1:}$ Los alumnos con mayor rendimiento son los que usan los servicios de la biblioteca por sobre el promedio.

Tabla 7-7: Prueba de proporciones entre promedios de notas y promedios de consultas

2012			
Promedios (sin uso cero)	Zscore <0,02	Zscore >=0,02	Total x bca
Uso bca >=23	42.94%	57.06%	4229
Uso bca<23	47.21%	52.79%	8803
Total x zscore	45.83%	54.17%	13032
X-squared = 20.805, df = 1, p-value = 5.086e-06			
2013			
Promedios (sin uso cero)	Zscore <0,01	Zscore >=0,01	Total x bca
Uso bca >=22	45.28%	54.72%	9594
Uso bca<22	47.66%	52.34%	20556
Total x zscore	46.91%	53.09%	30150
X-squared = 10.16, df = 1, p-value = 0.001435			
2014			
Promedios (sin uso cero)	Zscore <0,02	Zscore >=0,02	Total x bca
Uso bca >=22	43.90%	56.10%	10816
Uso bca<22	47.06%	52.94%	22820
Total x zscore	46.05%	53.95%	33636
X-squared = 29.48, df = 1, p-value = 5.65e-08			

Se observa que, durante los tres años, los usuarios con notas por sobre el promedio que consultan en biblioteca son cerca de un 54% del total. De estos, quienes usan la biblioteca por sobre el promedio, corresponden a alrededor del 56%.

La mayoría de los usuarios usan la biblioteca por debajo del promedio, y de estos el 52% tiene notas sobre el promedio.

El chi cuadrado que se obtiene siempre es más alto que el valor crítico, por lo que se deduce que existe una dependencia entre las variables.

El p-value para los tres años es <a 0.05, por lo que la diferencia se considera significativa, y se rechaza la hipótesis nula, se acepta la hipótesis de que los alumnos con mejor rendimiento consultan más la biblioteca. Se continuará analizando más en detalle esta relación más adelante.

7.1.7 Notas y multidisciplinariedad.

Con el fin de analizar si existe alguna relación entre la multidisciplinariedad de las consultas y los resultados académicos de los alumnos, se comparó para los años 2012, 2013 y 2014, la cantidad de áreas en que los estudiantes pidieron material y las notas obtenidas.

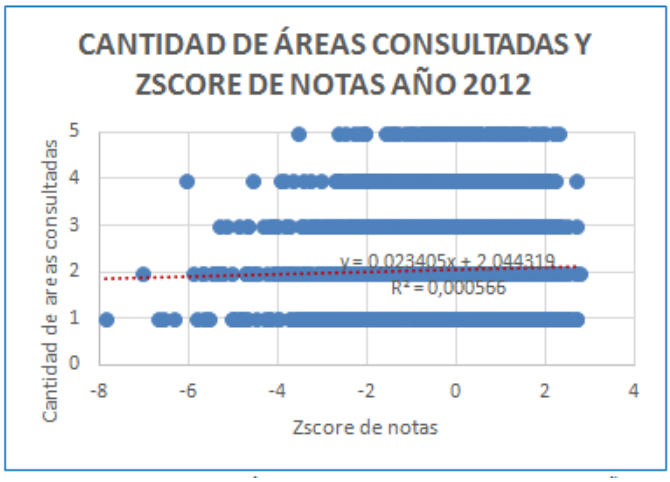

Imagen 7—17: Cantidad de áreas consultadas y zscore de notas año 2012

Se observa en la imagen anterior, que al considerar el total de alumnos que consultaron la colección presencial durante el 2012, las notas obtenidas y la cantidad de consultas realizadas, la distribución de las notas para las consultas desde 1 a 5 áreas es similar, con una área de distribución de notas compartida, la que abarca principalmente el zscore de notas entre -2 y 2. La diferencia se observa principalmente como una disminución de las

colas de la distribución, en la medida que aumenta la cantidad de áreas consultadas tienden a disminuir los estudiantes con zscore menor y levemente los con zscore mayor.

El índice de correlación ajustado R^2 es muy bajo, un 0.000566, lo que indica que hay una relación lineal casi nula entre notas y cantidad de áreas consultadas.

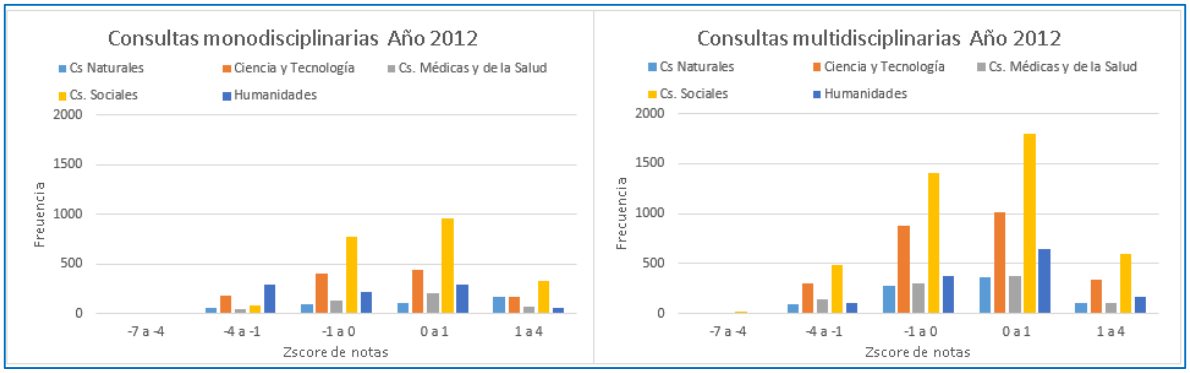

Imagen 7—18: Multidisciplinariedad de las consultas por área del usuario y notas año 2012

Al considerar el área del usuario para analizar la relación entre notas y cantidad de áreas consultadas, para el 2012 los alumnos del área de Ciencias Sociales y Ciencia y tecnología, que consultaron tanto sólo un área y 2 áreas y más, presentaron el mayor zscore de notas, Humanidades destaca por consultas multidisciplinarias más que de una sola área, como se observa en la imagen 7-16.

Se realiza el mismo análisis para el año 2013, al considerar el total de transacciones presenciales de los alumnos, sus notas y la cantidad de áreas consultadas, se aprecia que las notas se mantienen similares para las consultas de 1 a 5 áreas, con un área de zscore común entre -2 y 2, al igual que el año anterior. Se observa una disminución de la densidad las consultas de alumnos con bajo zscore y un poco más leve en aquellos con el más alto. Como se puede observar en la siguiente imagen.

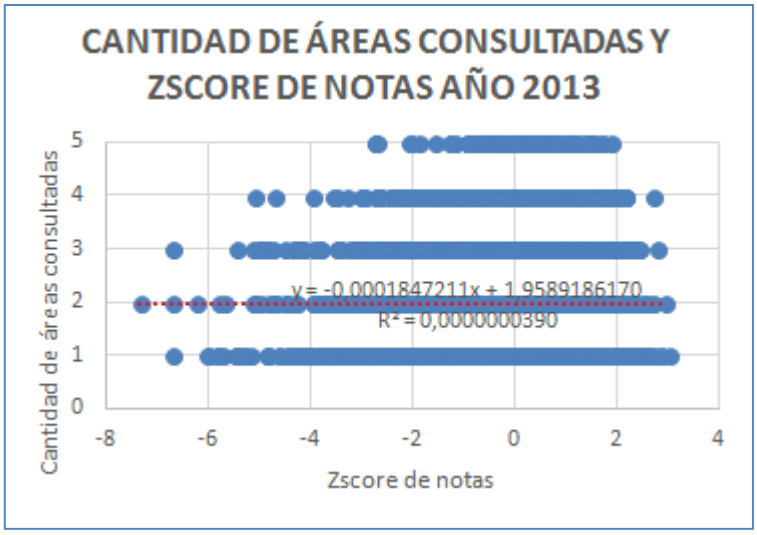

Imagen 7—19: Cantidad de áreas consultadas y zscore de notas año 2013

El índice de correlación ajustado R^2 es cercano a 0, un = 0.0000000390, menor que el año anterior, lo que indica que casi no existe una relación lineal entre notas y cantidad de áreas consultadas.

Al realizar el análisis considerando las áreas de los usuarios, la cantidad de áreas consultadas y las notas, para el 2013 observamos que los alumnos de las áreas Ciencia y Tecnología y Ciencias Sociales son quienes presentan el mejor Zscore de notas, entre -1 y 1, consultando más de 1 área y también exclusivamente 1 área. Los alumnos de Humanidades con el mejor Zscore de notas han consultado 3 y 4 áreas y los de Ciencias Naturales 3 áreas. Respecto al año 2012, en el área de Cs. Sociales los alumnos con zscore entre 0 y 1 generaron un aumento de sus consultas monodisciplinarias, se aprecia en la siguiente imagen.

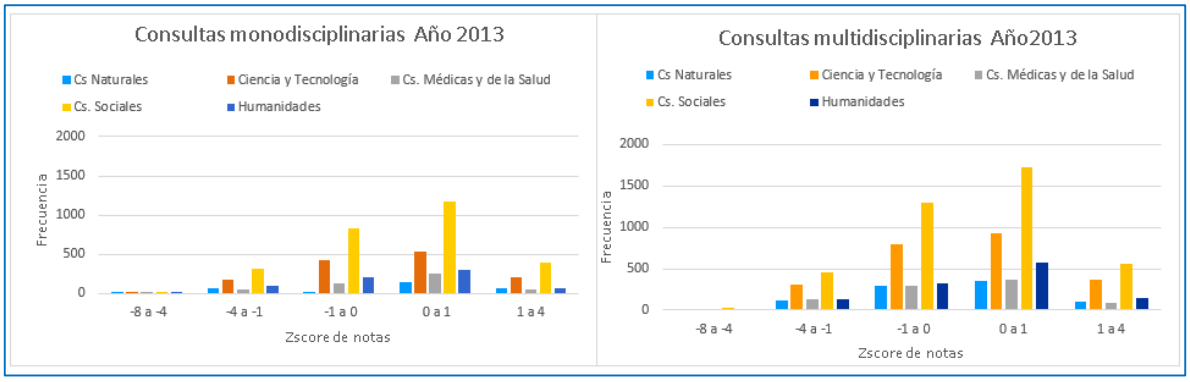

Imagen 7—20: Multidisciplinariedad de consultas por área del usuario y notas año 2013

Para el año 2014, al analizar la cantidad de transacciones respecto a las notas obtenidas y la cantidad de áreas consultadas, se observa que en comparación a los años anteriores hay una leve disminución de las consultas en 5 áreas y la densidad se mantiene alta entre los zscore -2 y 2 y baja en las puntas (ver imagen 7-19).

El índice de correlación ajustado R^2 es cercano a 0, un = 0.000102, mayor que el año anterior y menor que el del 2012, lo que indica que la relación lineal entre notas y cantidad de áreas consultadas casi no existe.

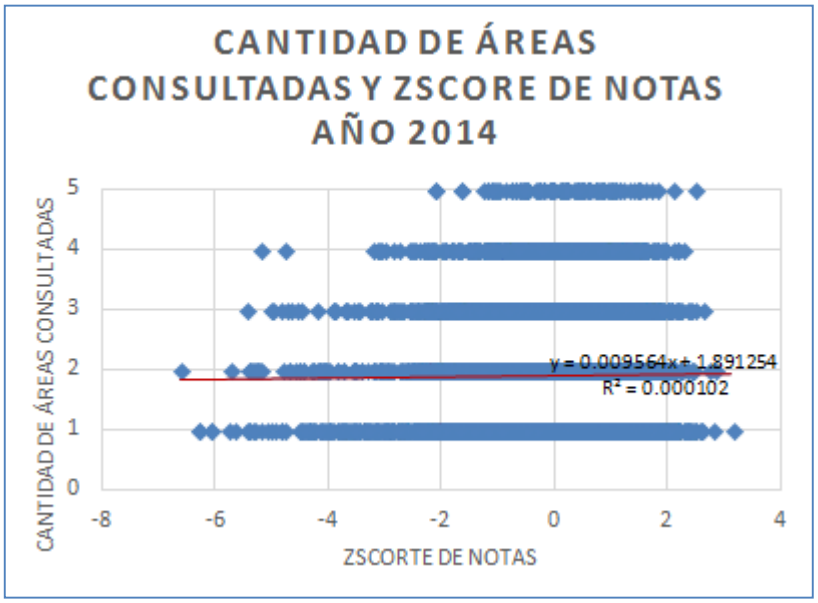

Imagen 7—21: Cantidad de áreas consultadas y zscore de notas año 2014

Al analizar las notas de los alumnos y la cantidad de áreas consultadas considerando el área de los alumnos, nos encontramos que para el 2014 las mejores notas en todas las áreas corresponden a consultas multidisciplinarias en 5 áreas, como se aprecia en la imagen 7-20. Para los alumnos de Ciencias Naturales las notas de quienes consultan 1 a 5 áreas se encuentran sobre zscore 0, en Ciencia y Tecnología y Humanidades quienes consultas de 3 a 5 áreas presentan las mejores notas (sobre zscore 0). En Ciencias Sociales quienes consultan 5, 3 y 2 áreas presentan las mejores notas. En Ciencias Médicas las mejores notas son de quienes consultan 5 y 1 área.

Al comparar los datos de 2014 con los años anteriores, se observa que los alumnos de Cs. Sociales y Ciencia y Tecnología presentaron un aumento de sus consultas monodisciplinarias, especialmente aquellos del rango de zscore entre -1 y 1. Para las consultas multidisciplinarias se mantiene la frecuencia en los distintos rangos de Zscore, como se aprecia en la siguiente imagen.

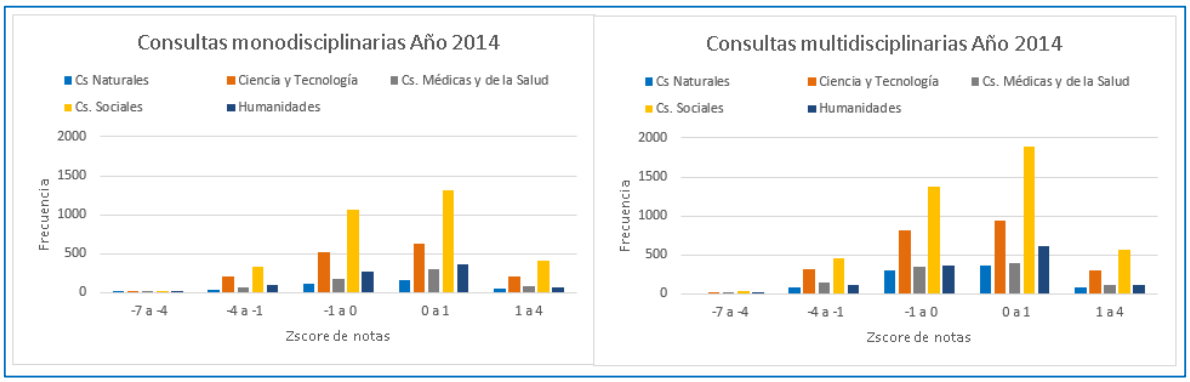

Imagen 7—22: Multidisciplinariedad de consultas por área del usuario y notas año 2014

Al revisar los datos compilados de los tres años, destacan los usuarios de Cs. Sociales y Ciencia y Tecnología, que tanto para las consultas monodisciplinarias y multidisciplinarias son quienes más consultan, en forma transversal a los rangos de zscore, tanto en los más bajos como en los más altos, estos alumnos presentan las mayores frecuencias, como se observa en la siguiente imagen. Además se presenta un mayor número de consultas en 2 áreas o más en todos los rangos de notas y en todas las áreas.

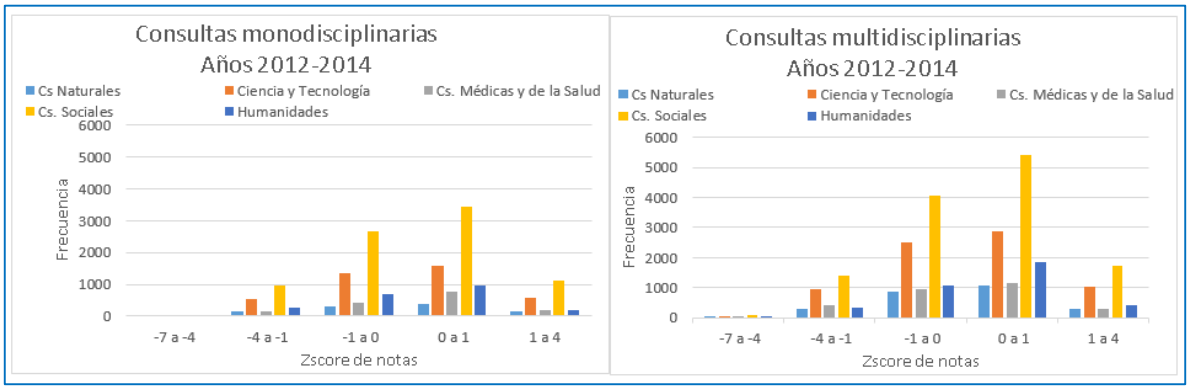

Imagen 7—23: Multidisciplinariedad de consultas por área del usuario y notas 2012-2014

Al realizar una revisión detallada de los datos compilados de los tres años, y separarlos por las áreas de los usuarios, se detectan patrones similares entre las áreas, como por ejemplo,

en todas se observa una mayor cantidad de consultas en 2 áreas y más, los rangos de zscore en los que existe el número de consultas más altos son -1 a 0 y 0 a 1, Ciencias Naturales y Ciencias Médicas tienen las frecuencias más bajas y Ciencias Sociales y Ciencia y Tecnología las más altas como se muestra en la siguiente imagen.

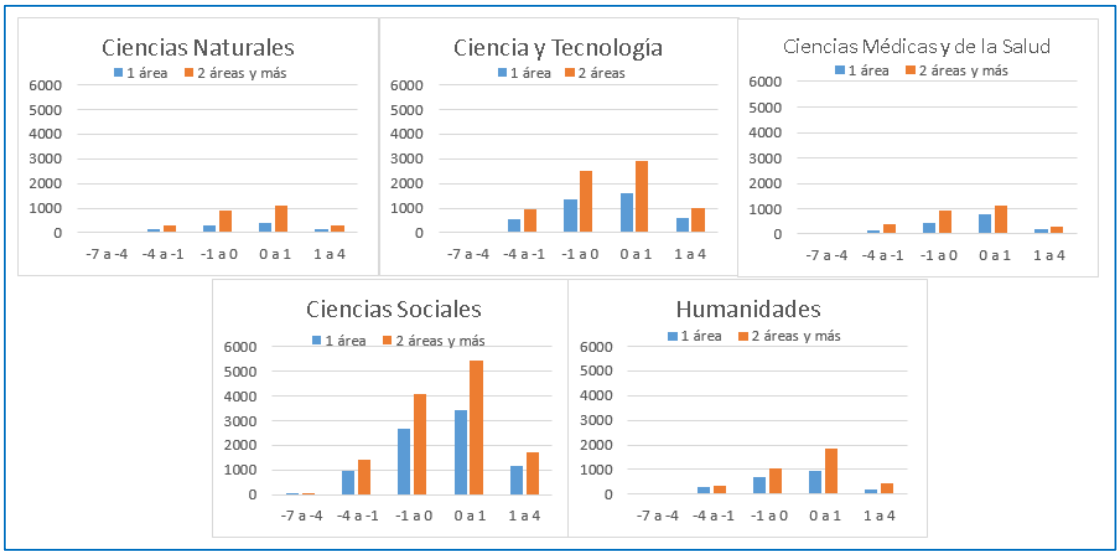

Imagen 7—24: Multidisciplinariedad y notas por áreas de los usuarios, años 2012-2014

Para continuar el análisis de multidisciplinariedad se procede a calcular las proporciones de las consultas de acuerdo a las notas y a la cantidad de áreas consultadas. En primer lugar se observa que las consultas en 1 y 2 áreas corresponden al 74,19% de las consultas realizadas por nuestros usuarios, siendo las consultas en 2 áreas levemente superior, y sólo un cuarto de las consultas corresponden a 3, 4 y 5 áreas. Si se considera el zscore de notas, el 74,57% se encuentra entre -1 y 1, siendo el rango 0-1 donde se encuentra la mayor cantidad de alumnos.

Tabla 7-8: Tabla de proporciones entre notas y cantidad de áreas consultadas

Zscore de notas	1 área	2 áreas	3 áreas	4 áreas	5 áreas	Total
-8 a -4	0.21%	0.17%	0.08%	0.01%	0.00%	0.47%
-4 a -1	4.61%	4.33%	2.13%	0.75%	0.08%	11.91%
-1-0	**11.82%**	**12.14%**	6.14%	1.98%	0.22%	32.32%
0-1	**15.48%**	**15.79%**	8.12%	2.58%	0.28%	42.25%
1 a 4	4.93%	4.70%	2.52%	0.78%	0.12%	13.05%
Total general	37.05%	37.14%	18.99%	6.11%	0.71%	100.00%

En la tabla anterior se observa que el grupo mayor de usuarios corresponde a quienes consultan 2 áreas y sus notas alcanzan al rango de zscore 0-1, luego siguen los que piden en 1 área y tienen el mismo zscore, después se encuentran los que piden en 2 áreas y tienen un zscore entre -1 y 0 seguidos por quienes consultan un área y tienen el mismo zscore. En estos cuatro grupos se concentra el 55.23% del total de los alumnos estudiados.

Posteriormente se incorporan las áreas del usuario para comparar las proporciones, se observa que el 45,31% de los usuarios que consultan provienen del área de Ciencias Sociales, y la mayoría de estos piden en 2 áreas o más y tienen un zscore entre 0 y 1. En la siguiente tabla se pueden ver los porcentajes de las otras áreas.

Tabla 7-9: Tabla de proporciones entre notas y cantidad de áreas consultadas por área del usuario

Áreas	**Cs. Naturales**		**Ciencia y Tecnología**		**Cs. Médicas**		**Cs. Sociales**		**Humanidades**		**Total**
Zscore notas	1 área	2 o más	1 área	2 o más	1 área	2 o más	1 área	2 o más	1 área	2 o más	
-8 a -4	0.02%	0.00%	0.05%	0.03%	0.01%	0.03%	0.09%	0.15%	0.03%	0.06%	0.47%
-4 a -1	0.35%	0.63%	1.21%	2.01%	0.37%	0.89%	2.08%	3.02%	0.60%	0.74%	11.91%
-1-0	0.68%	1.91%	2.92%	5.42%	0.96%	2.03%	5.76%	8.84%	1.50%	2.31%	32.32%
0-1	**0.87%**	**2.34%**	**3.46**%	**6.26%**	**1.64%**	**2.47%**	**7.44%**	**11.74%**	**2.07%**	**3.96%**	42.25%
1 a 4	0.35%	0.65%	1.26%	2.19%	0.45%	0.64%	2.46%	3.71%	0.40%	0.92%	13.05%
Total general	2.27%	5.54%	8.89%	15.91%	3.43%	6.05%	17.85%	27.46%	4.60%	8.00%	100.00%

Para continuar el análisis de la multidisciplinariedad de las consultas, se calculó el coeficiente de correlación Pearson considerando como variables el zscore de notas y la cantidad de áreas consultadas, para los años 2012, 2013 y 2014 por el área de las carreras de los alumnos. Se observa en la siguiente imagen los índices de correlación obtenidos considerando las áreas de los usuarios.

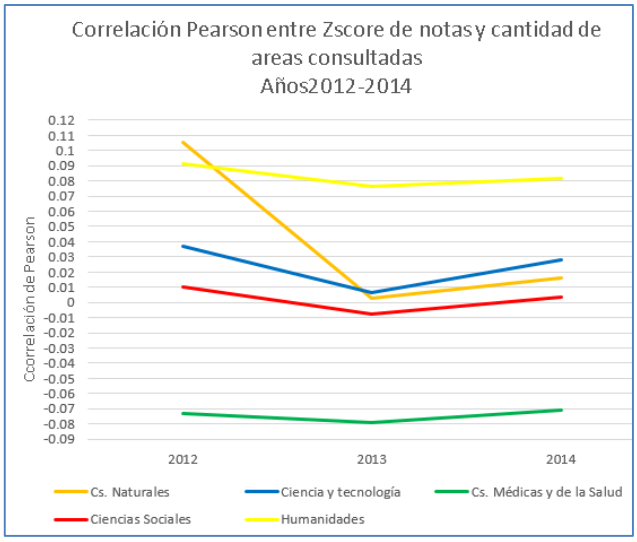

Imagen 7—25: Correlación entre Zscore de notas y cantidad de áreas consultadas años 2012-2014

Las áreas que presentan el indicador de correlación más alto en los tres años, son Humanidades, Ciencias Naturales y Ciencia y Tecnología, seguidos por Ciencias Sociales, con la correlación más baja se encuentra Ciencias Médicas y de la Salud. Al ser un índice de correlación cercano a 0 no es muy significativa la asociación entre las variables.

En Cs. Naturales se observa un fuerte descenso en la correlación para el año 2013, que luego sube levemente el 2014. En Ciencias Médicas y de la Salud se observa un indicador negativo, que va subiendo levemente en los años de estudio. En todas las áreas se observa un leve aumento en la correlación durante el 2014.

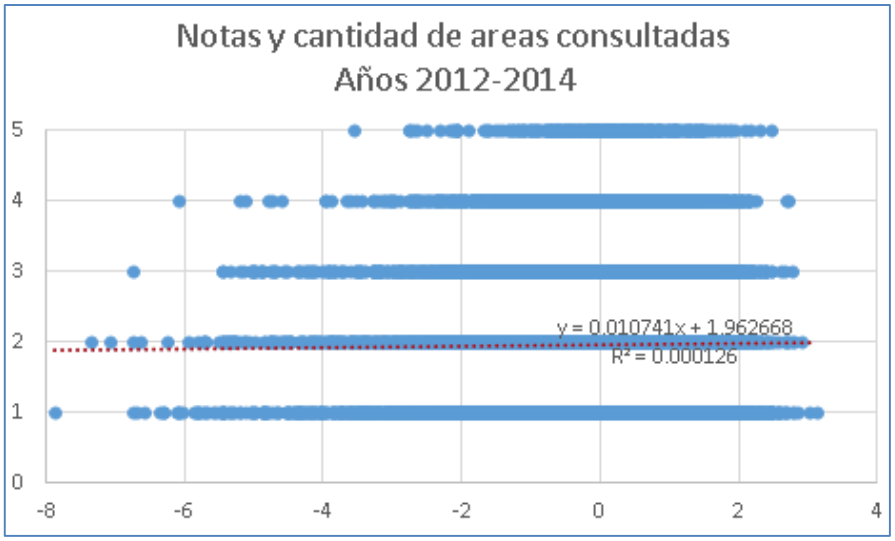

Imagen 7—26: Notas y cantidad de áreas consultadas durante los años 2012 y 2014

En general se aprecia que existe una leve, casi nula, relación entre las variables de cantidad de áreas consultadas y las notas obtenidas. Se podría decir entonces que la multidisciplinariedad de las consultas no se encuentra asociada en forma positiva con las notas obtenidas por los alumnos.

Para realizar un análisis estadístico lo más completo posible y confirmar si existe o no relación entre estas variables se extraen los siguientes indicadores a los datos compilados 2012-2014. En primer lugar nos planteamos la pregunta de investigación ¿existe relación entre la multidisciplinariedad de las consultas y el rendimiento académico de los alumnos?

Consideramos como variable principal el Zscore de notas y como variable secundaria Cantidad de áreas consultadas. La hipótesis nula es H_0= No hay relación entre las variables Notas y cantidad de áreas consultadas. La hipótesis alternativa es H_1= Hay relación entre las variables Notas y cantidad de áreas consultadas.

Se realiza un análisis descriptivo sobre los datos de Zscore de los alumnos y la cantidad de áreas que consultaron durante los años 2012 y 2014, los valores obtenidos se pueden observar en la siguiente tabla.

Tabla 7-10: Análisis estadístico de notas y cantidad de áreas consultadas

Estadísticas descriptivas	*Zscore notas*	*Cantidad de áreas*
Media	0,020443002	1,962887624
Error típico	0,004530007	0,004331987
Mediana	0,116995	2
Moda	0,563565	2
Desviación estándar	0,973803997	0,931236206
Varianza de la muestra	0,948294225	0,867200872
Curtosis	3,602375941	-0,084092486
Coeficiente de asimetría	-1,114710435	0,738045601
Rango	11,007426	4
Mínimo	-7,868242	1
Máximo	3,139184	5
Suma	944,691563	90707
Cuenta	46211	46211
Mayor (1)	3,139184	5
Menor(1)	-7,868242	1
Nivel de confianza (95,0%)	0,008878883	0,008490761

Luego se determina el nivel de dependencia de las variables con un análisis de regresión lineal, cuyos resultados se observan a continuación.

Tabla 7-11: Regresión lineal de notas y áreas consultadas por los alumnos

Estadísticas de la regresión	Valores
Coeficiente de correlación múltiple R	0,011232
Coeficiente de determinación R^2	0,000126
R^2 ajustado	0,000105
Error típico	0,973753
Observaciones	46211

$R^{2,}$ o tamaño del efecto de una variable sobre otra, es 0.000126, es decir solo conseguiremos explicar el 0.0126% de las notas obtenidas, mediante su dependencia con la cantidad de áreas consultadas.

Luego se realiza la prueba estadística t para evaluar si dos grupos difieren entre sí de manera significativa respecto a sus promedios aritméticos.

Tabla 7-12: Prueba t de notas y áreas consultadas por los alumnos

Prueba t	Zscore notas	Cantidad de áreas
Media	0,020443002	1,962887624
Varianza	0,948294225	0,867200872
Observaciones	46211	46211
Coeficiente de correlación de Pearson	0,011232001	
Diferencia hipotética de las medias	0	
Grados de libertad	46210	
Estadístico t	-311,6550197	
P(T<=t) una cola	0	
Valor crítico de t (una cola)	1,644886603	
P(T<=t) dos colas	0	
Valor crítico de t (dos colas)	1,960015266	

La probabilidad asociada al estadístico t es inferior a 0.05, por lo que se advierte una diferencia significativa en las variables.

Para calcular el chi cuadrado, que prueba la independencia de dos variables entre sí, y el p-value, que nos indica si existe diferencia significativa con la hipótesis nula, se realiza la

prueba de proporciones, con las variables cantidad de áreas consultadas y notas de los alumnos, de cada año por separado. Los resultados se observan en la siguiente tabla.

Se puede apreciar que mientras la mayoría de los alumnos consultan 2 áreas y más, entre un 33% y 40% consultan material de un área. El 55% de quienes consultan, tanto un área como dos o más, tienen notas sobre el promedio.

El porcentaje de quienes tienen notas sobre el promedio y consultan un área ha aumentado y decrecido, quienes consultan dos áreas o más y tienen notas sobre el promedio se mantienen en el 55%.

Tabla 7-13: Prueba de proporciones para cantidad de áreas consultadas y notas

2012			
Promedios	**Zscore <0**	**Zscore >=0**	**Total x bca**
1 área	46.42%	53.58%	4989
2 áreas y más	44.58%	55.42%	9944
Total x zscore	45.20%	54.80%	14933
X-squared = 4.4792, df = 1, p-value = 0.03431			
2013			
Promedios	**Zscore <0**	**Zscore >=0**	**Total x bca**
1 área	43.52%	56.48%	5595
2 áreas y más	44.42%	55.58%	9422
Total x zscore	44.08%	55.92%	15017
X-squared = 1.108, df = 1, p-value = 0.2925			
2014			
Promedios	**Zscore <0**	**Zscore >=0**	**Total x bca**
1 área	44.96%	55.04%	6537
2 áreas y más	44.72%	55.28%	9724
Total x zscore	44.82%	55.18%	16261
X-squared = 0.078094, df = 1, p-value = 0.7799			

El chi cuadrado es mayor al valor crítico (3.84) sólo para el año 2012, para 2013 y 2014 tiene valores menores, lo que indica que existe una independencia entre las variables.

El p-value es < a 0.05 sólo para el año 2012, 2013 y 2014 presentan valores mayores o que nos indica que la diferencia respecto a la hipótesis nula es significativa sólo para el año

2012, para el año 2013 y 2014 no es significativa. Se observa una tendencia a aumentar el p-value y disminuir el chi cuadrado en los años estudiados.

Se acepta la hipótesis nula y queda invalidada entonces la hipótesis alternativa que dice que hay relación entre las variables Notas y cantidad de áreas consultadas, aunque el 2012 esta haya sido muy débil y con una significación muy baja. Por esta razón no se realizarán posteriores análisis en esta área de la multidisplinariedad.

7.1.8 Clustering de transacciones y notas

Para determinar formalmente la existencia de grupos de usuarios con características comunes, se procede a aplicar algoritmos de clustering sobre los set de datos.

El objetivo de aplicar algoritmos de clustering es buscar grupos de usuarios con características similares entre ellos y con diferencias con los otros grupos, considerando las variables notas y cantidad de transacciones.

Se representan los resultados en las matrices de dispersión obtenidas para cada algoritmo aplicado por año, considerando las dos variables en estudio.

Se trabajó con un subconjunto del set de datos, compuesto por el ID del usuario, las transacciones presenciales realizadas y su Zscore para los años 2012, 2013 y 2014. Este set de datos se analizó en Rapid Miner con los algoritmos de Kmeans y de DBscan, realizando pruebas con distintos indicadores, obteniendo los resultados que se detallan a continuación.

7.1.8.1 K-means

Este algoritmo es exclusivo, es decir cada usuario es asignado a uno de los conjuntos de clúster indicados en K. Los usuarios en cada conjunto son similares entre ellos, esta similitud se basa en la medida de las distancias entre ellos, en un clúster se encuentran aquellos que están más cercanos. Se seleccionó un K=4, el algoritmo generó los siguientes grafos:

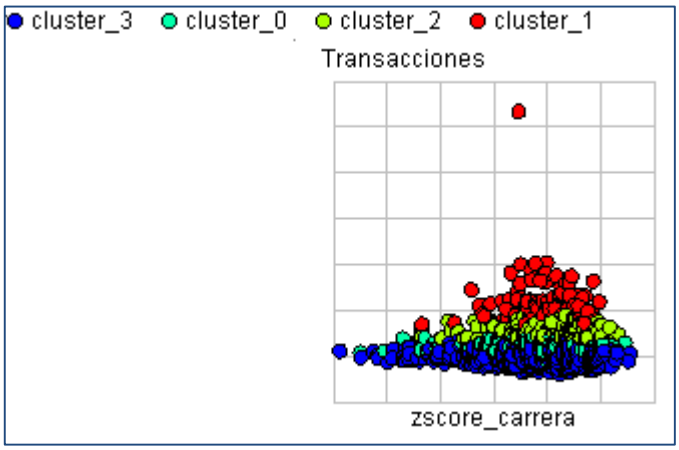

Imagen 7—27: Clusters de transacciones y Zscore con K-means para 2012

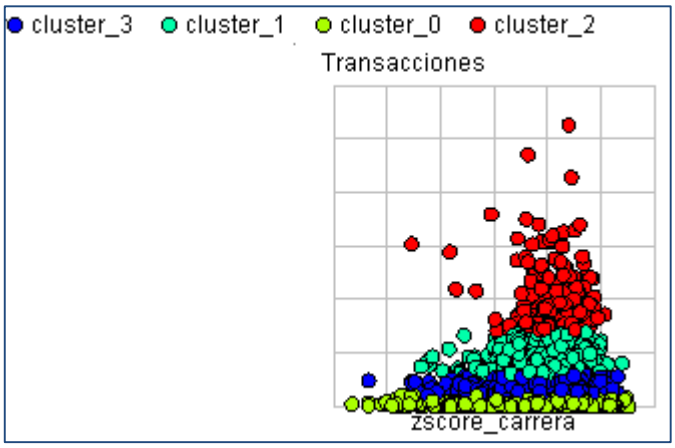

Imagen 7—28: Clusters de transacciones y Zscore con K-means para 2013

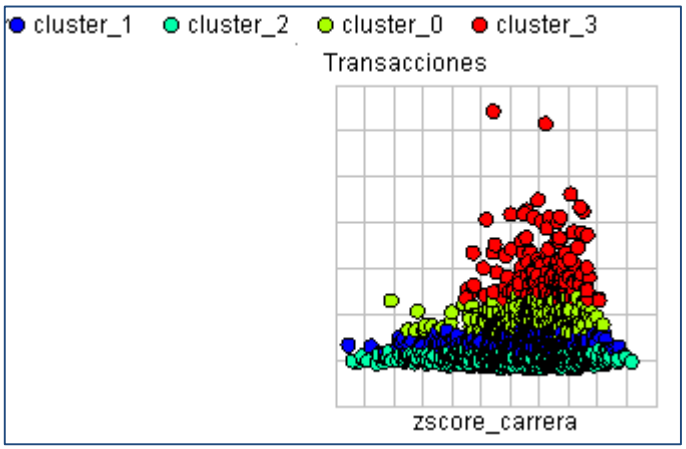

Imagen 7—29: Clusters de transacciones y Zscore con K-means para 2014

Se observa cierta tendencia a la concentración de la densidad en quienes tienen menos transacciones y una dispersión en los grupos con mayor cantidad de transacciones.

La distribución entre los rangos de Zscore es similar, observándose que quienes tienen menos transacciones se distribuyen en todos los rangos de Zscore. Quienes tienen más transacciones se encuentran entre quienes tienden a tener un Zscore mayor. Quienes tienen un bajo Zscore tienden a tener bajas transacciones

A través de los años se nota un aumento en la distribución de las transacciones, el algoritmo fija los cuatro grupos por cantidad de transacciones y en los años 2013 y 2014 se aprecia la aparición de varios casos de alto Zscore y alto número de transacciones.

Tabla 7-14: Vectores de desempeño de los clusters con K-means

Vector de desempeño de los clusters	**2012**	**2013**	**2014**
Promedio de distancia intercluster	-62.875	-57.828	-61.257
Promedio de distancia en clúster 0	-46.000	-67.596	-22.998
Promedio de distancia en clúster 1	-11.407	-25.678	-40.324
Promedio de distancia en clúster 2	-27.831	-12.926	-71.617
Promedio de distancia en clúster 3	-73.569	-41.987	-11.169

Se observa que las distancias entre los clúster se mantienen similares en los tres años estudiados, mientras que las distancias de los puntos dentro de los clúster se dispersan notoriamente en los años. El sistema identifica con clúster distintos los diferentes sectores de la dispersión de los datos, por eso en la tabla hemos mantenido los colores de los clúster para identificar cada sector del gráfico.

En la base del gráfico, aquellos alumnos con pocas transacciones y diversos Zscore, desde el más bajo al más alto (identificados el 2012 con clúster 3 azul, el 2013 son el clúster 0 verde y el 2014 clúster 2 calipso) son los que mantienen las mayores distancias o dispersión entre los puntos, entre -67 y -73 mil.

El segundo grupo identificado, aquellos que tiene un poco más de transacciones y no presentan los Zscore más bajos (identificados el 2012 con clúster 0 calipso, el 2013 son el clúster 3 azul y el 2014 clúster 1 azul), mantienen distancias muy similares, alrededor de los -40 mil.

El tercer grupo identificado, aquellos alumnos con mayor cantidad de transacciones y con un Zscore cercano al promedio y por sobre el promedio (identificados el 2012 con clúster 2 verde, el 2013 son el clúster 0 calipso y el 2014 clúster 0 verde), mantienen distancias muy similares, alrededor de los -25 mil.

El cuarto grupo identificado, aquellos alumnos con la mayor cantidad de transacciones y con un Zscore por sobre el promedio (identificados el 2012 con clúster 1 rojo, el 2013 con el clúster 2 rojo y el 2014 con clúster 3 rojo), mantienen distancias muy similares, alrededor de los -12 mil.

7.1.8.2 DBSCAN

Este operador es un algoritmo de clustering basado en la densidad, ya que encuentra agrupaciones considerando la distribución de densidad estimada de los nodos a partir de un punto inicial. Se utiliza un Epsilon de 5.0 y un punto mínimo de 5.

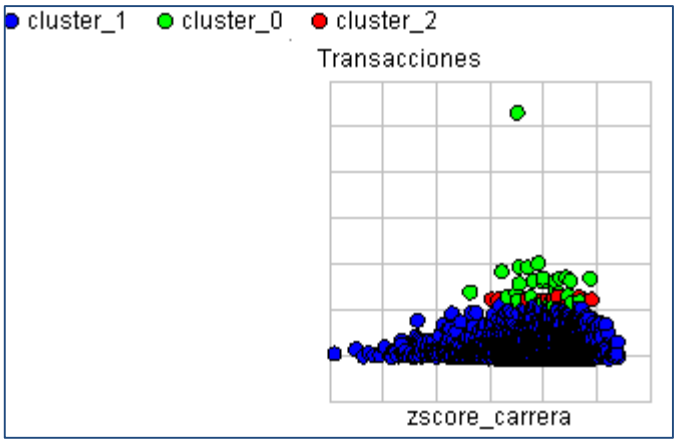

Imagen 7—30: Clusters de transacciones y Zscore con DBSCAN para 2012

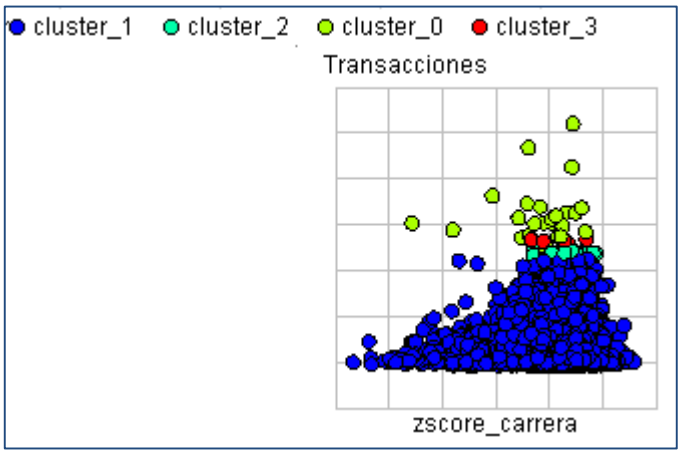

Imagen 7—31: Clusters de transacciones y Zscore con DBSCAN para 2013

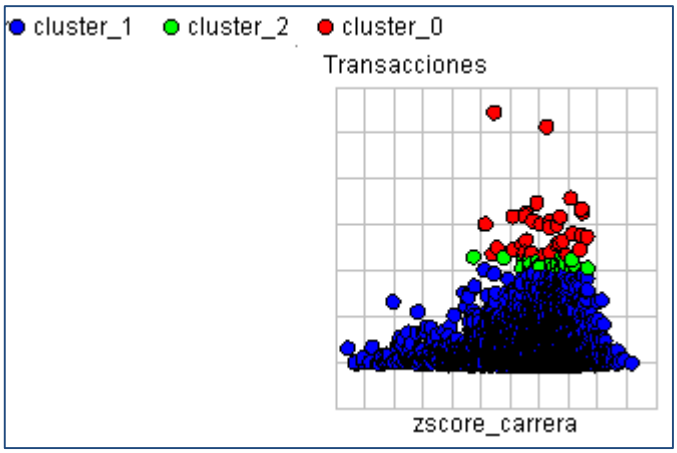

Imagen 7—32: Clusters de transacciones y Zscore con DBSCAN para 2014

El algoritmo en los distintos años, con los mismos indicadores, generó 3 y 4 clusters. Se aprecia una alta densidad en el primer cluster, identificando que quienes tienen menos transacciones, están por debajo de la media, se distribuyen por todos los rangos del Zscore, al igual que en kmeans, es factible identificar que los usuarios con altas transacciones tienen alto Zscore, y quienes tienen bajo Zscore tienen bajas transacciones. No existen usuarios con altas transacciones y bajo Zscore.

Tabla 7-15: Vector de desempeño de los clusters con DBSCAN

Vector de desempeño de los clúster	2012	2013	2014
Promedio de distancia intercluster	**-409.631**	**-387804**	**-364143**
Promedio de distancia en clúster 0	**-3.368**	**-1479**	**-2128**
Promedio de distancia en clúster 1	**-410807**	**-389123**	**-365464**
Promedio de distancia en clúster 2	**-54**	**-135**	**-230**
Promedio de distancia en clúster 3		**-19**	

El primer grupo compuesto por quienes tienen menos transacciones, por debajo de la media, y un Zscore disperso, es el que tiene las mayores distancias durante los años (grupo azul). Son los más numerosos. El segundo grupo, compuesto por alumnos con mejores notas y más consultas, posee una distancia similar en todos los años. Son menos que el clúster anterior. El tercer grupo (y cuarto del 2013), presenta una fuerte disminución en la densidad, y en las distancias entre los puntos. Es el grupo minoritario.

7.2 Consultas a la colección electrónica

Para realizar esta parte del análisis se consultó la base de datos creada por el proyecto FONDECYT 1131109, específicamente Matriz de rendimiento académico y transacciones 2012-2014, desde donde se generaron otros archivos con distintas variables.

7.2.1 Análisis estadístico

Las consultas a la colección electrónica provienen de usuarios registrados, que principalmente consultan desde fuera de la Red UC, y los usuarios no registrados, que consultan dentro de la Red UC y no requieren registrarse. La cantidad de usuarios registrados ha aumentado 2,56 veces (256%) desde el 2012 al 2014, mientras que los usuarios no registrados han ido disminuyendo.

Tabla 7-16: Cantidad de usuarios de la colección electrónica

Usuarios EZproxy	2012	2013	2014	Total
De Usuarios no registrados	108	176	92	376
De Usuarios registrados	18208	42117	46696	107021
Total general	18316	42293	46788	107397

La mayoría de los usuarios presentan varias consultas o sesiones durante el año, las consultas en general han aumentado 21,6 veces. Se calculó el promedio por año y se observa que en general el promedio de consultas por usuario para los tres años es de casi 27. Desde el 2012 al 2014 el promedio de consultas ha aumentado 8 veces (847%). Este aumento se observa fuertemente en los usuarios registrados y en menor medida con los no registrados, como se indica en la siguiente tabla.

Tabla 7-17: Cantidad de sesiones de consulta de la colección electrónica

Consultas EZproxy	2012	2013	2014	Total
Promedio por usuario	6,96	14,64	59,01	26,87
De Usuarios no registrados	2353	755	5314	8422
De Usuarios registrados	125232	618632	2755727	3499591
Total de consultas por año.	127585	619387	2761041	3508013

Ya hemos encontrado que los usuarios realizan en promedio casi 27 consultas al año, es interesante conocer la duración de estas sesiones, al analizar esta variable se observa que en promedio para los tres años las sesiones duran 177 minutos, equivalentes a 2, 95 horas. Es decir los usuarios pasan en promedio casi tres horas al año consultando la colección electrónica. Al analizar por año se observa que la duración de las sesiones en promedio ha aumentado 4,6 veces (465%), desde el 2012 al 2014, de 75, 68 minutos a 352 minutos por usuario, como lo presenta la tabla a continuación.

Tabla 7-18: Duración de las sesiones de consultas a la colección electrónica

Duración de consultas EZproxy (min.)	2012	2013	2014	Total
Promedio por usuario al año	75,68	104,06	352,27	177,33
Promedio de Usuarios no registrados	119,78	63,48	286,61	156,62
Promedio de Usuarios registrados	75,42	104,22	352,314	177,32

Los usuarios no registrados presentan una menor duración de las sesiones de consultas que los usuarios registrados, especialmente el año 2013 se observa una baja considerable en la duración de las sesiones de usuarios no registrados.

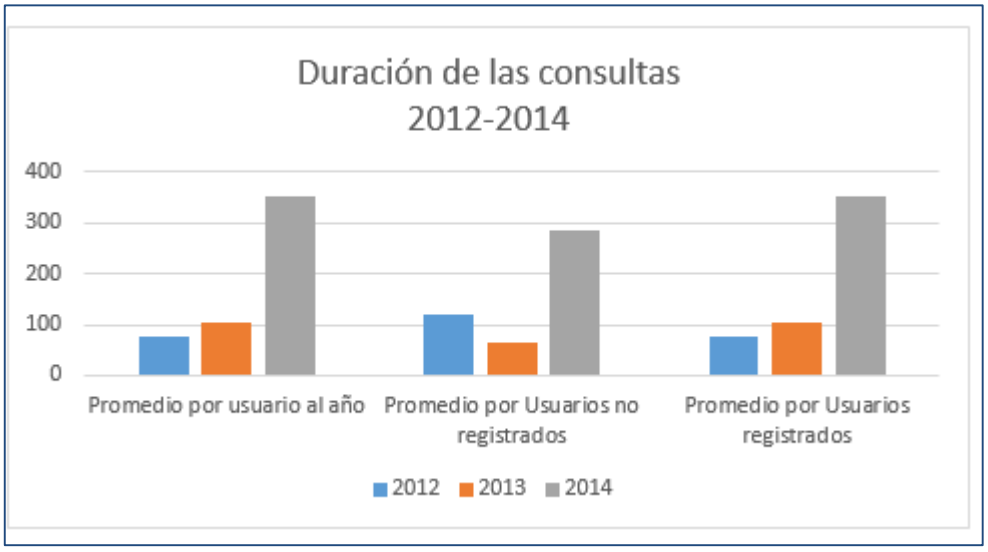

Imagen 7—33: Promedio de duración de consultas a la colección electrónica

7.2.2 Consultas por área y carrera del usuario

Para conocer los patrones de consultas de nuestros usuarios de la colección electrónica, se buscó identificar una relación entre las consultas y las áreas y carreras a las que pertenecen los usuarios.

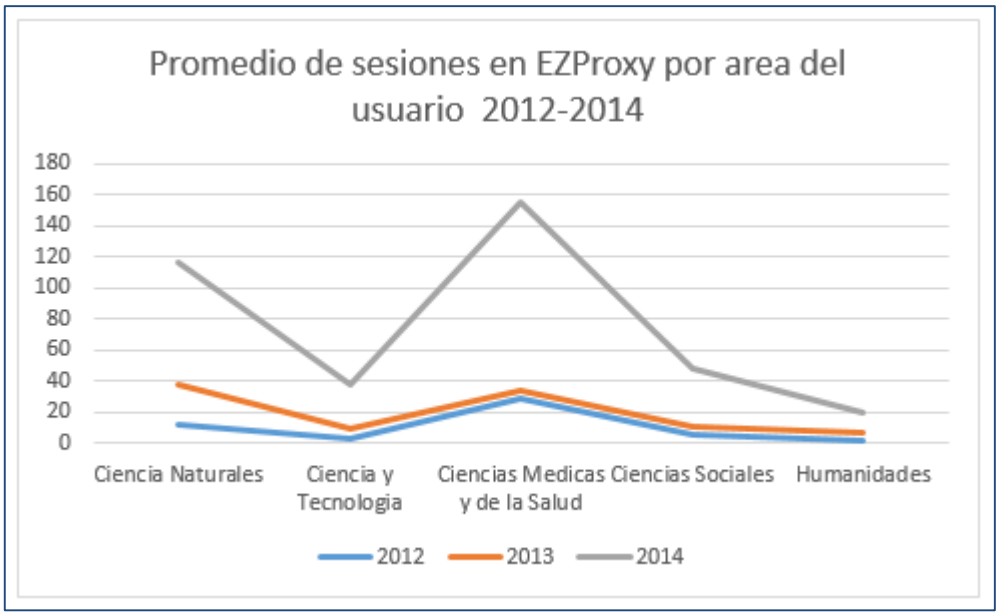

Imagen 7—34: Promedio de sesiones en EZproxy por área del usuario

Al contrastar las consultas con las áreas de los usuarios se observa que quienes tienen más consultas son los de Ciencias Médicas y de la Salud, tendencia que se mantiene en los años y se eleva drásticamente el año 2014. Los usuarios de Ciencias Naturales son los que presentan el segundo nivel de consultas, el que también se eleva el año 2014. Los usuarios de Humanidades, Ciencias Sociales y Ciencia y Tecnología presentan un bajo nivel de consultas, que presenta un aumento el año 2014.

Para determinar la relevancia de las consultas se considera también la duración de cada consulta por el área de los usuarios.

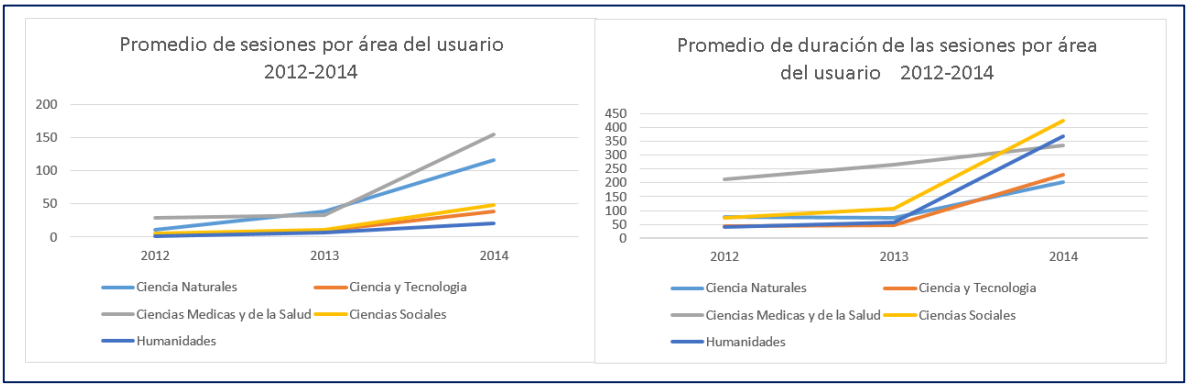

Imagen 7—35: Promedio de sesiones y duración de las sesiones por área del usuario

Se observa que los usuarios de Ciencias Médicas y de la Salud junto con tener el nivel más alto de consultas también tienen la duración más alta por consulta, los usuarios de Ciencias Sociales y Humanidades pese a tener un bajo nivel de consultas presentan una duración más alta que Ciencias Naturales y Ciencia y Tecnología.

En general los usuarios divididos por área tienen un promedio de 35 consultas con una duración de 175 minutos para los tres años, esto es 4,85 minutos de duración por sesión.

La distribución promedio de los minutos por sesión de acuerdo al área del usuario se ve en la siguiente tabla:

Tabla 7-19: Promedio de minutos por sesión y por área de los usuarios

Área	Min. por sesión
Ciencia Naturales	2,11
Ciencia y Tecnología	6,24
Ciencias Médicas y de la salud	3,76
Ciencias Sociales	9,34
Humanidades	16,40
Total general	4,85

Se observa que los usuarios que presentan el promedio más alto de duración por sesión son los del área de Humanidades, con menos sesiones al año, comparada con Cs. Naturales, pero de mayor duración.

En general, se observa un fuerte incremento de la cantidad de sesiones y su duración, en la siguiente tabla se presenta en detalle el porcentaje de crecimiento de estas e las distintas áreas de los usuarios desde el año 2012 al 2014.

Tabla 7-20: % de crecimiento de las consultas y su duración 2012-2014

Area del usuario	Consultas	Duración
Ciencia Naturales	1000,3	267,7
Ciencia y Tecnologia	1159,4	502,8
Ciencias Medicas y de la Salu	543,1	156,4
Ciencias Sociales	940,3	579,3
Humanidades	1096,8	924,1
Total general	847,2	465,5

Las áreas con un fuerte crecimiento de las consultas son Humanidades, Ciencia y Tecnología y Cs. Naturales. La con menor crecimiento es Cs. Médicas y de la Salud, tanto en cantidad de sesiones como en duración, pero aun así sigue manteniéndose como el área con mayor cantidad de consultas de la Universidad.

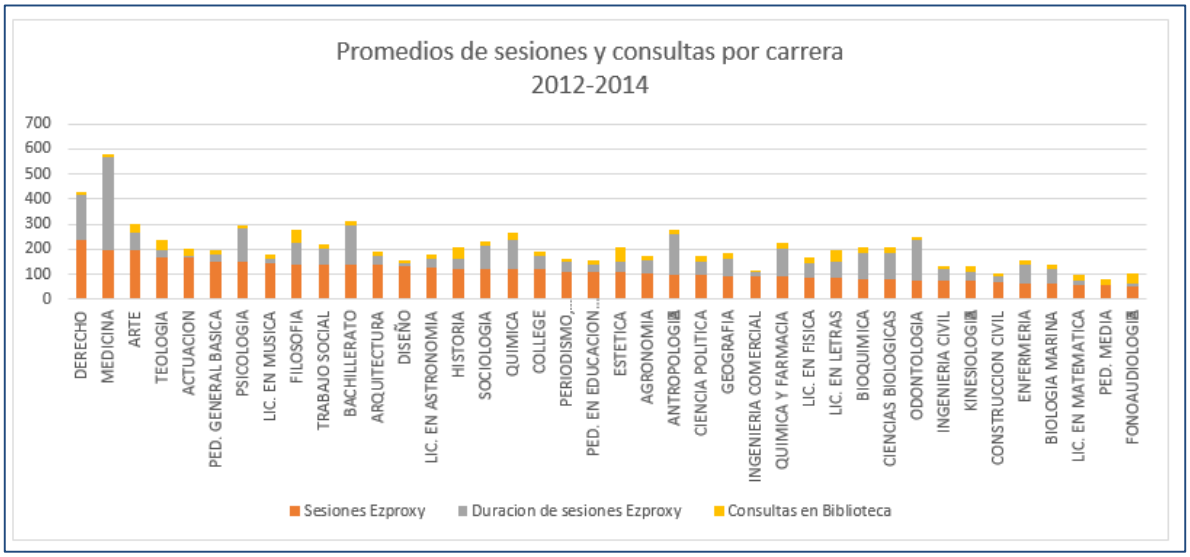

Imagen 7—36: Promedio de sesiones y consultas por carrera

A un nivel más específico, considerando las carreras de los usuarios, las consultas a la colección electrónica, la duración de estas sesiones y las consultas en bibliotecas, destacan las carreras de Derecho, Medicina, Arte y Teología, que presentan los mayores promedios en la cantidad de sesiones; Medicina, Derecho, Antropología y Odontología que presentan los mayores promedios de duración de las sesiones.. Las consultas en biblioteca presencial corresponden en general a un 23% de las sesiones a la colección electrónica.

Por carrera, el promedio de sesiones en la colección electrónica es de 111, con 72,32 minutos de duración, 22 consultas promedio en biblioteca y un promedio de duración de las sesiones a la colección electrónica de 1,53 minutos por sesión.

7.2.3 Consultas a la colección electrónica y uso de la biblioteca

Se considera luego el año de carrera como variable comparativa con las sesiones de consultas y el uso de material presencial en biblioteca. El material presencial en biblioteca se divide en impresos, audiovisual y uso de espacios (salas de estudio).

La cantidad de sesiones EZProxy respecto a las consultas presenciales se pueden observar por año de carrera en la siguiente figura.

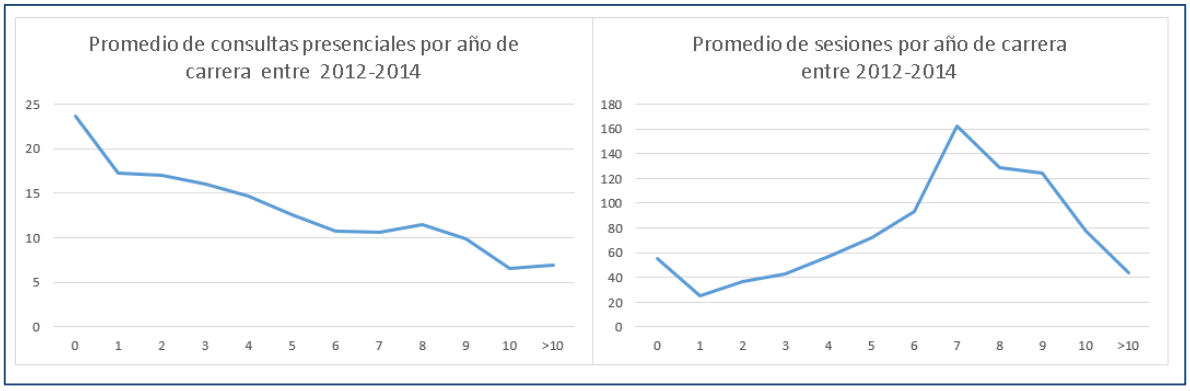

Imagen 7—37: Consultas presenciales y sesiones por año de la carrera

En la figura 7-26 se observa que la cantidad de sesiones Ezproxy aumentan en la medida que aumentan los años de estudio, con la excepción del primer año en que se aprecia un descenso, luego la curva asciende hasta el año 7, cuando comienza el descenso hasta el año 10, último considerado.

El uso de material impreso es mayor los primeros años de estudio y disminuye a partir del año 3. El uso de material audiovisual y de espacio, aunque es menor que el uso de impreso, mantiene una tendencia similar, es mayor los primeros años y disminuye a partir del tercer año.

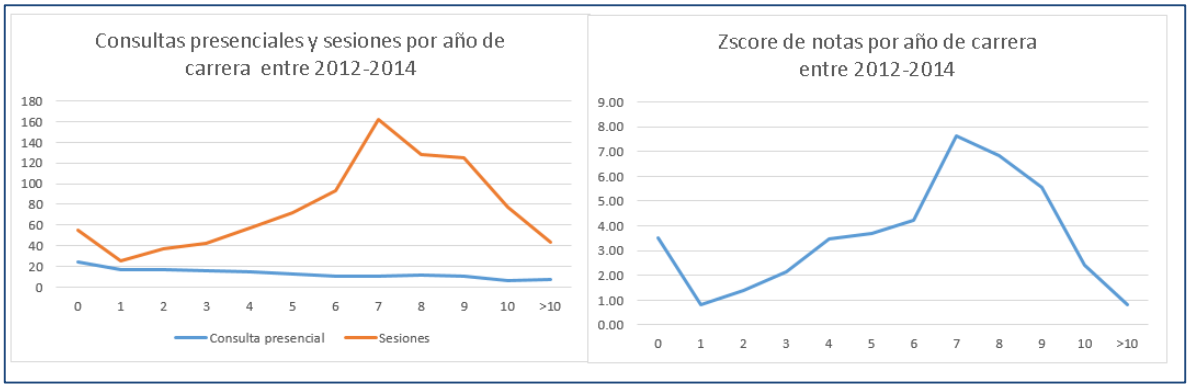

Imagen 7—38: Sesiones y Zscore de notas por año de carrera

Al comparar notas con sesiones EZproxy, su duración y el uso de biblioteca en los años 2012-2014, se observa en la imagen 7-27 que los alumnos que presentan una mayor duración de sesiones poseen notas cercanas al Zscore 0 y superior. Quienes presentan una mayor cantidad de sesiones presentan un Zscore de notas superior a 0, y quienes tienen un mayor uso de la biblioteca tienen notas cercanas a 0 o superiores.

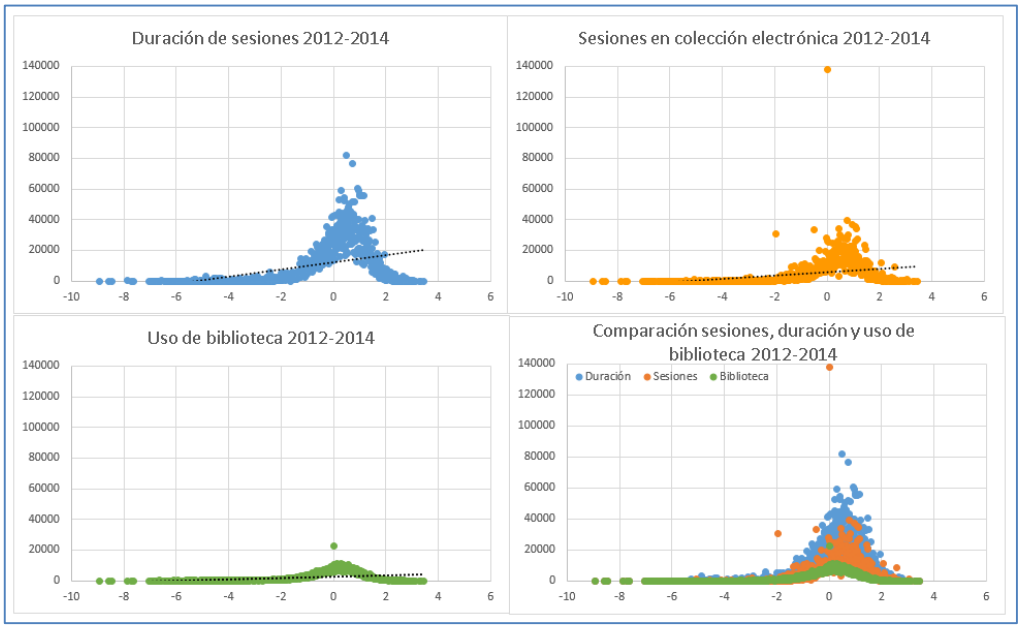

Imagen 7—39: Promedio de consultas por Zscore

Al comparar los promedio de consultas, duración y uso de la biblioteca por zscore, se observa que las tres variables tienen su máxima densidad alrededor del Zscore de notas 0, con una tendencia hacia el zscore sobre 1.

Se busca una correlación entre las variables antes mencionadas y se obtiene los valores mostrados en la tabla siguiente.

Tabla 7-21: Correlación entre variables Zscore, sesiones, duración y uso de biblioteca

Estadísticas de la regresión	*Sesiones-Uso de biblioteca*	Notas-Uso de biblioteca	Notas-Sesiones	Notas-Duración
Coeficiente de correlación múltiple	0.00121	0.03287	0.04827	0.09464
Coeficiente de determinación R^2	0.00000	0.00108	0.00233	0.00896
R^2 ajustado	-0.00001	0.00107	0.00232	0.00895

Después de aplicar un análisis de regresión, se observa que todas las variables analizadas en duplas tienen una correlación positiva, siendo la asociación entre notas y duración la más alta. Al calcular el R^2 la dependencia entre las variables baja un poco, y el indicador ajustado deja en negativo la asociación entre sesiones y uso de la biblioteca.

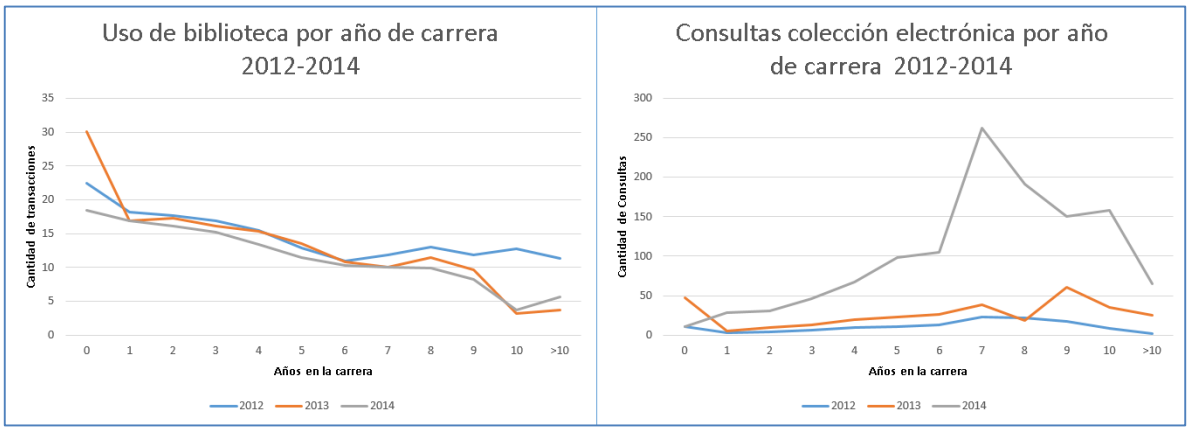

Imagen 7—40: Transacciones en biblioteca y consultas a Col. Electrónica por año de carrera

Finalmente se compara el promedio de transacciones presenciales en biblioteca y las consultas de la colección electrónica por año de carrera se observa que en la medida que aumentan los años de carrera va disminuyendo la cantidad de transacciones en biblioteca, esto es evidente para los años 2013 y 2014, el año 2012 no se observa una baja tan considerable los últimos años de carrera. En contraste, la consulta de la colección electrónica presenta un aumento en la medida que aumentan los años de carrera, los años 2012 y 2013 se mantiene un aumento moderado, el año 2014 es considerablemente alto a partir del tercer año, alcanzando el peak el año 7.

Sería erróneo establecer que las consultas a la colección electrónica aumentan en detrimento de las consultas presenciales, sino más bien se podría deducir que se van complementando, ya que el descenso de las consultas presenciales comienza el primer año de carrera y cae en el punto más bajo el año 10, y las consultas electrónicas aumentan a partir del tercer año y caen también el año 10. No se puede establecer una causalidad entre ambas variables.

Se analiza estadísticamente los valores obtenidos para las variables cantidad de sesiones y uso de biblioteca de los alumnos entre los años 2012 y 2014, en la tabla a continuación se ven los resultados de quienes consultaron biblioteca y/o generaron sesiones en la colección electrónica.

Tabla 7-22: Análisis estadístico de Sesiones y uso de la biblioteca 2012 a 2014

Análisis estadístico	*Suma de sesiones*	*Análisis estadístico2*	*Suma de Biblioteca*
Media	121,6995062	Media	58,90502852
Error típico	4,672820607	Error típico	0,590545132
Mediana	0	Mediana	26
Moda	0	Moda	0
Desviación estándar	792,3979548	Desviación estándar	100,1422469
Varianza de la muestra	627894,5187	Varianza de la muestra	10028,46961
Curtosis	1651,982526	Curtosis	83,29740304
Coeficiente de asimetría	31,09100843	Coeficiente de asimetría	6,037674458
Rango	59798	Rango	3210
Mínimo	0	Mínimo	0
Máximo	59798	Máximo	3210
Suma	3499591	Suma	1693873
Cuenta	28756	Cuenta	28756
Mayor (1)	59798	Mayor (1)	3210
Menor(1)	0	Menor(1)	0
Nivel de confianza(95,0%)	9,158945355	Nivel de confianza(95,0%)	1,157495879

Luego se realiza un análisis de la correlación de las variables en cada año, con el tamaño del efecto, y el ajuste de este. Los resultados se observan en la siguiente tabla:

Tabla 7-23: Correlación para cantidad de sesiones y uso de biblioteca

Estadísticas	***2012***	***2013***	***2014***
Coeficiente de correlación	0.0191	0.0380	0.0418
Tamaño del efecto	0.0004	0.0014	0.0017
R^2 ajustado	0.0003	0.0014	0.0017
Error típico	23.07	47.44	54.96
Observaciones	28756	28756	28756

El coeficiente de correlación para los tres años es leve, entre 0.01 y 0.04, el tamaño del efecto y su ajuste es siempre cercano a cero. A pesar de obtener resultados tan bajos, se observa un leve aumento de estos índices en los tres años estudiados, teniendo los resultados más altos para el año 2014.

Luego se analizan los datos en conjunto para los 3 años, en el siguiente gráfico podemos observar la dispersión de las variables y la línea de tendencia con el tamaño del efecto.

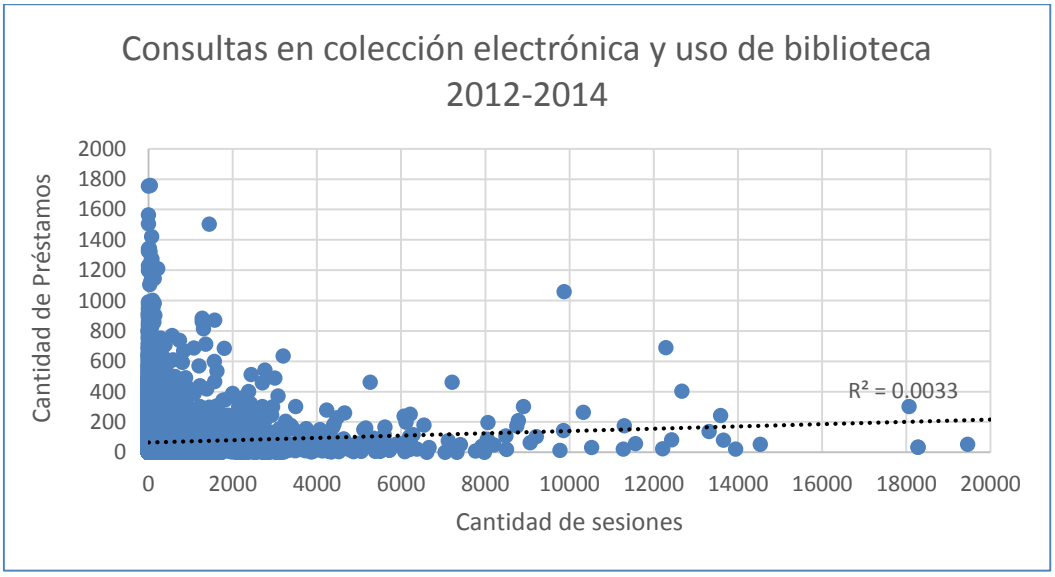

Imagen 7—41: consultas en la colección electrónica y uso de la biblioteca

Se observa una mayor densidad entre quienes usan poco la biblioteca y tienen pocas sesiones en la colección electrónica. Se ven pocos casos con alta cantidad de préstamos y sesiones. En la medida que aumentan en una variable, disminuyen en la otra. El tamaño del efecto el cercano a cero, 0.0033, lo que junto a la línea de tendencia, indica una leve relación, casi inexistente.

De acuerdo a estos resultados, si se considera las siguientes hipótesis laterales al estudio:

H_0: No existe relación entre las consultas de la colección electrónica y el uso de la biblioteca presencial

H_1: Existe relación entre las consultas de la colección electrónica y el uso de la biblioteca presencial.

Se puede indicar que existe una relación entre las variables de uso de biblioteca y cantidad de sesiones, pero esta es muy leve, aunque ha ido creciendo en los años.

7.2.4 Consultas a la colección electrónica y notas

Para buscar la relación entre cantidad de sesiones realizadas y notas obtenidas por los alumnos se analiza un set de datos con el zscore de notas y la cantidad de sesiones realizadas por los alumnos, durante los 3 años de estudio. Se comienza graficándolas para cada uno de los años.

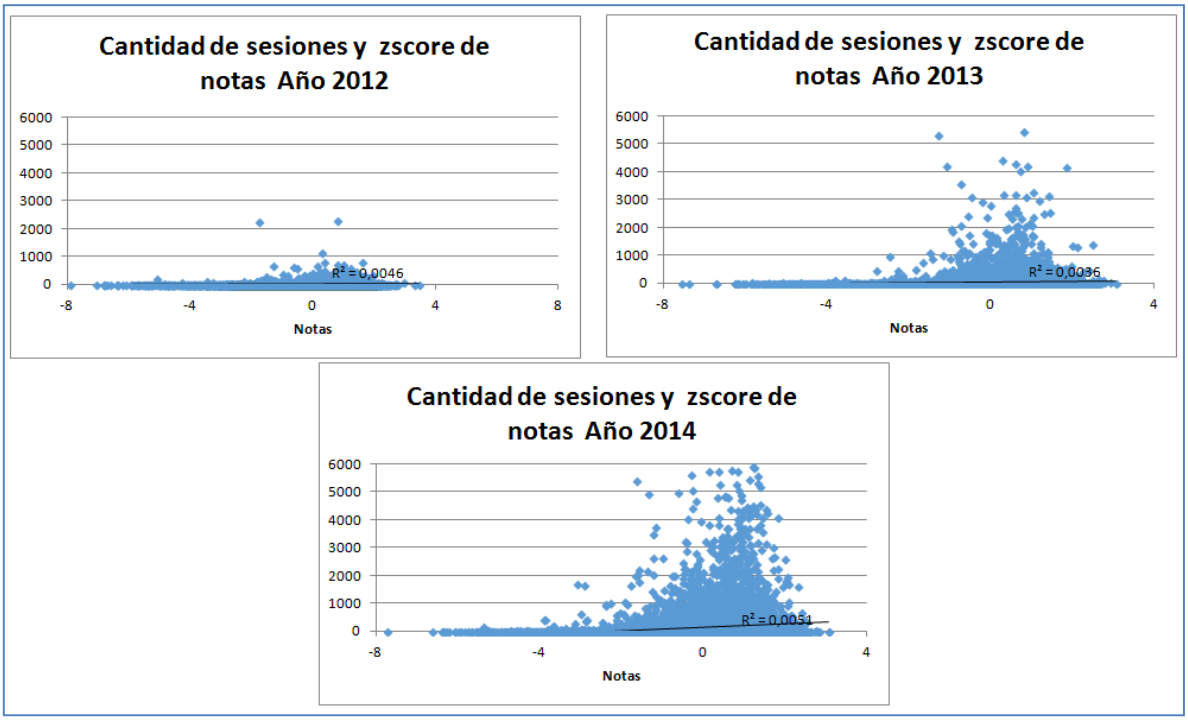

Imagen 7—42: Cantidad de sesiones y Zscore de notas obtenidas por año

Se observa que la cantidad de sesiones y la distribución de las notas han ido en aumento con los años, el coeficiente de correlación indica que existe una asociación cercana a 0 de las variables. Se observa una tendencia a una mayor densidad en las áreas cercanas a Zscore 0.

Al analizar el conjunto de cantidad de sesiones promedio por zscore de notas de los alumnos observamos que el mayor promedio de sesiones se encuentra entre los zscore 1 y 2, como se observa en el gráfico a continuación.

Imagen 7—43: Cantidad de sesiones promedio por zscore de notas.

Posteriormente se genera una regresión lineal, en la siguiente tabla se encuentran algunos de los resultados.

Tabla 7-24: Regresión lineal para sesiones y notas

Estadísticas de la regresión	*2012*	2013	2014
Coeficiente de correlación mú	0,0680552	0,059997279	0,071359757
Coeficiente de determinación	0,00463151	0,003599674	0,005092215
R^2 ajustado	0,00457683	0,003552982	0,005046725
Error típico	0,99580428	0,960193782	0,93807432
Observaciones	18206	21342	21873

La correlación se mantiene cercana a cero en todo los años, disminuye al calcular R2 ajustado, pero siempre es positiva.

Luego se realiza un test de proporciones para cada año, comparando el promedio anual de la cantidad de sesiones y el promedio anual de Zscore de notas. Los resultados se encuentran en la siguiente tabla.

Las hipótesis a analizar son:

$H_{o:}$ No existe diferencia entre el rendimiento académico de los alumnos que usan los servicios de la biblioteca por sobre o bajo el promedio.

$H_{1:}$ Los alumnos con mayor rendimiento son los que usan los servicios de la biblioteca por sobre el promedio.

Tabla 7-25: Prueba de proporciones entre Cantidad de sesiones y Zscore de notas

2012			
Promedios	**Zscore <0.2**	**Zscore >=0.2**	**Total**
Sesiones >=31	35.07%	64.93%	978
Sesiones <31	46.46%	53.54%	3125
Total x zscore	43.75%	56.25%	4103
X-squared = 38.823, df = 1, p-value = 4.64e-10			
2013			
Promedios	**Zscore <0.2**	**Zscore >=0.2**	**Total**
Sesiones >114	35.10%	64.90%	1057
Sesiones <114	45.72%	54.28%	4368
Total x zscore	43.65%	56.35%	5425
X-squared = 38.59, df = 1, p-value = 5.229e-10			
2014			
Promedios	**Zscore <0.2**	**Zscore >=0.2**	**Total**
Sesiones >=398	35.28%	64.72%	1437
Sesiones <398	48.31%	51.69%	5483
Total x zscore	45.61%	54.39%	6920
X-squared = 77.415, df = 1, p-value < 2.2e-16			

Se observa que para cada año cerca del 55% de quienes consultan tienen notas iguales o sobre el promedio anual del Zscore de notas. De estos alumnos más de un 64% tiene consultas por sobre el promedio anual. Los alumnos con zscore por debajo del promedio tienen un 44% de las consultas, de estos la mayoría tienen sesiones por debajo del promedio anual.

En cuanto a la cantidad de sesiones, aproximadamente el 78% de los alumnos que consultan lo hacen por debajo del promedio de consultas anuales. Cerca del 22% presenta sesiones por sobre el promedio, de estos, la mayoría, tiene notas por sobre el promedio.

El promedio de consultas ha ido aumentando rápidamente al igual que la cantidad de usuarios que consulta la colección electrónica. En todos los años se observa la tendencia de que la mayoría de los usuarios con notas sobre el promedio consultan la colección electrónica sobre el promedio.

El chi cuadrado obtenido se muestra con valores por sobre el valor crítico (3,86), lo que nos indica que existe una dependencia de las variables.

El p-value obtenido es 2.2e-16, menor a 0.05, lo que indica que existe una diferencia significativa entre las variables como para rechazar la hipótesis nula. Se acepta la hipótesis de que los alumnos con mayor rendimiento son los que consultan más los servicios de la biblioteca.

Se realizó el mismo análisis para las variables duración de las sesiones y notas de los alumnos, en la siguiente tabla se encuentra el test o prueba de proporciones,

Tabla 7-26: Prueba de proporciones entre duración de las sesiones y notas de los alumnos

2012			
Promedios	**Zscore <0.2**	**Zscore>=0.2**	**Total**
Duración >=335	39,77%	60,23%	1222
Duración <335	45,44%	54,56%	2881
Total x zscore	43,75%	56,25%	4103
X-squared = 10.96, df = 1, p-value = 0.0009312			
2013			
Promedios	**Zscore <0.2**	**Zscore>=0.2**	**Total**
Duración >=452	38,73%	61,27%	1575
Duración <452	45,66%	54,34%	3850
Total x zscore	43,65%	56,35%	5425
X-squared = 21.557, df = 1, p-value = 3.435e-06			
2014			
Promedios	**Zscore <0.2**	**Zscore>=0.2**	**Total**
Duración >523	39,90%	60,10%	2080
Duración <523	48,06%	51,94%	4840
Total x zscore	45,61%	54,39%	6920
X-squared = 38.664, df = 1, p-value = 5.035e-10			

Se observa en los tres años que la mayoría de los usuarios (cerca del 70%) tiene consultas cortas, bajo el promedio de duración. De estos el 53% aproximadamente tiene notas sobre el promedio.

Quienes tienen consultas sobre el promedio de duración corresponden a cerca del 30%, de estos más del 60% tiene notas sobre el promedio.

El chi cuadrado es más alto que el valor crítico (3,84), por lo que se aprecia una dependencia entre las variables. El p-value es en todos los años menor a 0.05, lo que indica una diferencia significativa entre las variables y que existe una asociación entre las variables.

7.2.5 Análisis de dominios consultados

Se agregó al análisis la variable de dominios para tener una idea de cuáles son los sitios más consultados a partir de los logs de EZproxy. Se considera sólo la raíz de las URLs de los sitios consultados, con el fin de medir las consultas a nivel de dominio y no de recurso específico, ya que esto sería entrar en el detalle por cada paper dentro de las bases de datos.

Se realiza una comparación entre la cantidad de consultas realizadas durante los años de estudio y la cantidad de dominios consultados. En la medida que la cantidad de consultas aumenta en los tres años, la cantidad de dominios consultados aumentó del 2012 al 2013 y disminuyó levemente el 2014. Con lo que se podría deducir que los usuarios están consultando menos dominios pero con mayor frecuencia.

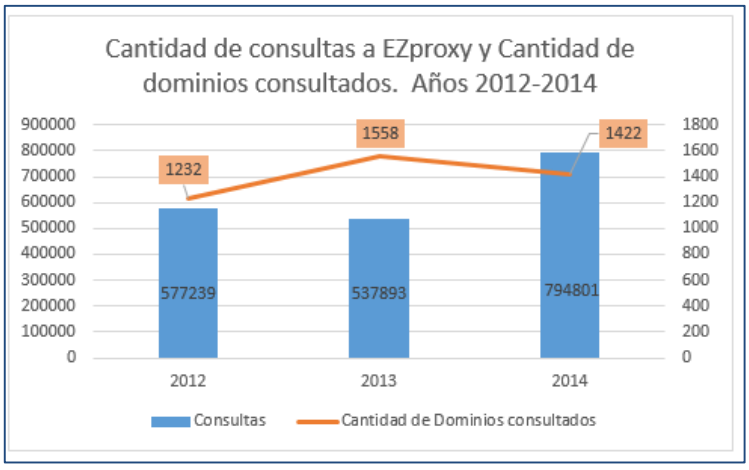

Imagen 7—44: Cantidad de consultas y dominios consultados

Se continua el análisis de los dominios creando una matriz con todos los dominios consultados durante los tres años, luego se observa aquellos dominios que de un año a otro no fueron consultados, aquellos que son nuevos y no fueron consultados anteriormente, los consultados durante los tres años y aquellos consultados sólo durante un año. Los resultados se indican en el gráfico siguiente.

Imagen 7—45: Consultas por dominios

El año 2012 muestra que tiene la mayor cantidad de dominios no consultados, esto sugiere que aparecieron después de ese año y/o fueron suscritos posteriormente, a su vez hubo 210 sitios que sólo fueron consultados ese año, los que podrían corresponder a recursos que fueron discontinuados o cambiaron de nombre.

El año 2013 tiene 608 sitios no consultados durante ese año, lo que indica que aún no eran suscritos o estaban discontinuados. Los sitios que fueron consultados sólo ese año fueron 304 y los consultados a partir de ese año fueron 573.

Para el año 2014 tenemos 744 sitios no consultados, que pueden corresponder a sitios discontinuados o con cambio de nombre. 361 recursos fueron consultados sólo este año, lo que indica que son recursos nuevos.

Existen 759 dominios que han sido consultados todos los años, al menos una vez por año, de estos dominios los diez con mayor cantidad de consultas durante los tres años son:

- www.sciencedirect.com
- www.ncbi.nlm.nih.gov
- search.proquest.com
- web.ebscohost.com
- www.uptodate.com
- media.proquest.com

- content.ebscohost.com
- onlinelibrary.wiley.com
- static.pubmed.gov
- apps.webofknowledge.com

En la siguiente figura se aprecia el detalle de los 50 dominios que recibieron más consultas durante los tres años y la cantidad de consultas por dominio para cada año.

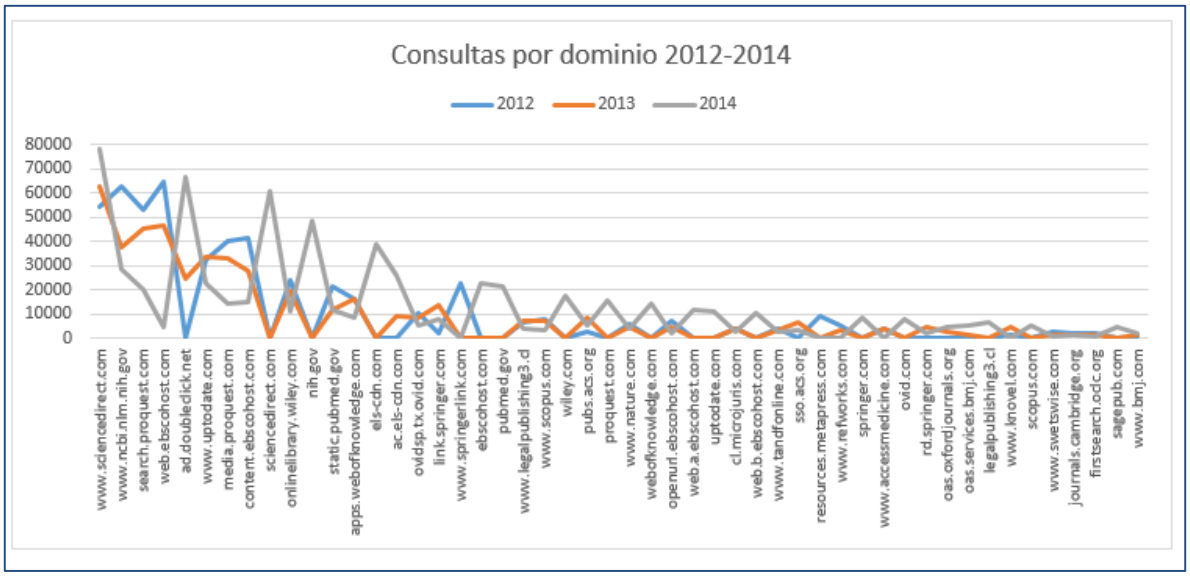

Imagen 7—46: Consultas por dominio en detalle

El año 2014 se aprecia un aumento notorio de las consultas en los distintos dominios, con algunas excepciones, como la de “web.ebscohost.com” cuyas consultas fueron disminuyendo del 2012 al 2014, seguramente porque el dominio fue cambiando paulatinamente a “ebscohost.com”.

Se observa que otros dominios cambian de nombre o se diversifican, pero se mantienen entre los más consultados, como Sciencedirect, Proquest, Springer, Wiley.

A partir de estos análisis se puede inferir que los usuarios de la colección electrónica mantienen un patrón estable de consulta en los dominios, o sus derivados. Consultan preferentemente recursos de dominios como Sciencedirect, Proquest, Springer y Wiley.

7.2.6 Conjuntos frecuentes de dominios consultados

Para continuar con un análisis profundo de los dominios más consultados, se utiliza el software Rapid Miner para determinar conjuntos frecuentes de dominios consultados en los tres años. Se utiliza el modelo de asociación y conjuntos frecuentes, específicamente el operador FP-growth. Considerando un soporte mayor a 0.15. Se obtiene listas de conjuntos frecuentes de dominios para cada año, se puede observar un extracto de los conjuntos obtenidos en la siguiente tabla.

Tabla 7-27: Conjunto frecuentes de dominios consultados

Soporte	2012	Soporte2	2013	Soporte3	2014
0,735	web.ebscohost.com - - -	0,665	web.ebscohost.com - -	0,528	sciencedirect.com - -
0,651	www.sciencedirect.com - - -	0,655	www.sciencedirect.com - -	0,469	ebscohost.com - -
0,627	content.ebscohost.com - - -	0,56	search.proquest.com - -	0,458	els-cdn.com - -
0,627	web.ebscohost.com - content.ebscohost.com - -	0,509	content.ebscohost.com - -	0,457	sciencedirect.com - els-cdn.com -
0,582	search.proquest.com - - -	0,509	web.ebscohost.com - content.ebscohost.com -	0,445	www.sciencedirect.com - -
0,512	web.ebscohost.com - www.sciencedirect.com - -	0,467	web.ebscohost.com - search.proquest.com -	0,388	proquest.com - -
0,502	web.ebscohost.com - search.proquest.com - -	0,465	web.ebscohost.com - www.sciencedirect.com -	0,383	ac.els-cdn.com - -
0,462	media.proquest.com - - -	0,434	media.proquest.com - -	0,383	www.sciencedirect.com - ac.els-cdn.com -
0,462	search.proquest.com - media.proquest.com - -	0,434	search.proquest.com - media.proquest.com -	0,326	sciencedirect.com - ebscohost.com -
0,439	www.sciencedirect.com - content.ebscohost.com - -	0,414	www.sciencedirect.com - search.proquest.com -	0,303	ebscohost.com - proquest.com -
0,439	content.ebscohost.com - search.proquest.com - -	0,383	link.springer.com - -	0,302	search.proquest.com - -
0,439	web.ebscohost.com - www.sciencedirect.com -	0,373	search.proquest.com - content.ebscohost.com -	0,293	content.ebscohost.com - -
0,439	web.ebscohost.com - content.ebscohost.com -	0,373	web.ebscohost.com - search.proquest.com - content.ebscohost.cc	0,292	sciencedirect.com - proquest.com -
0,426	www.sciencedirect.com - search.proquest.com - -	0,366	web.ebscohost.com - media.proquest.com -	0,288	ebscohost.com - els-cdn.com -
0,402	web.ebscohost.com - media.proquest.com - -	0,366	web.ebscohost.com - search.proquest.com - media.proquest.com	0,288	sciencedirect.com - ebscohost.com - els-cdn.com
0,402	web.ebscohost.com - search.proquest.com - media.proquest.com -	0,362	www.sciencedirect.com - content.ebscohost.com -	0,287	sciencedirect.com - www.sciencedirect.com -
0,38	www.springerlink.com - - -	0,362	web.ebscohost.com - www.sciencedirect.com - content.ebscohos	0,284	web.a.ebscohost.com - -
0,379	web.ebscohost.com - www.sciencedirect.com -	0,356	web.ebscohost.com - www.sciencedirect.com - search.proquest.c	0,275	web.b.ebscohost.com - -
0,358	content.ebscohost.com - media.proquest.com - -	0,348	www.sciencedirect.com - link.springer.com -	0,271	springer.com - -
0,358	web.ebscohost.com - content.ebscohost.com -	0,347	onlinelibrary.wiley.com - -	0,261	els-cdn.com - proquest.com -
0,358	content.ebscohost.com - search.proquest.com -	0,322	www.sciencedirect.com - media.proquest.com -	0,261	sciencedirect.com - els-cdn.com - proquest.com
0,358	web.ebscohost.com - content.ebscohost.com - search.proquest.com - media.proquest.com	0,322	www.sciencedirect.com - search.proquest.com - media.proquest.	0,259	els-cdn.com - www.sciencedirect.com -
0,351	openurl.ebscohost.com - - -	0,312	www.sciencedirect.com - onlinelibrary.wiley.com -	0,259	sciencedirect.com - els-cdn.com - www.sciencedirect.com
0,346	www.sciencedirect.com - www.springerlink.com - -	0,303	content.ebscohost.com - media.proquest.com -	0,257	sciencedirect.com - ac.els-cdn.com -

En los tres años estudiados los conjuntos frecuentes con mayor soporte, es decir aquellos que fueron consultados con mayor periodicidad, incluyen recursos de Ebscohost, Sciencedirect, Proquest y Springer.

Se observa también aquí que los dominios van cambiando de nombre o derivan nuevos sub-dominios, pero en general se mantienen como base los nombres de los dominios mencionados anteriormente.

El mayor soporte de los tres años, de 0,735, se presentó el año 2012 con el dominio web.ebscohost.com, el que bajó a 0.665 el 2013 y a 0.469 el año 2014.

El año 2013 incluye también el domino onlinelibrarywiley.com, con un soporte 0,347, que no había aparecido en los más frecuentes del 2012.

El año 2014 aparecen nuevos dominios entre los con mayor soporte, como el caso de els-cdn.com, con un soporte de 0,383.

Se observa que en la medida que pasan los años los soportes máximos son menores que el año anterior, lo que se podría explicar al considerar que ha ido aumentando la cantidad de bases de datos y recursos electrónicos que se incluyen en la colección cada año y las mismas bases de datos se van diversificando y generan nuevos dominios más específicos, por lo tanto las consultas también se van distribuyendo en los distintos dominios. Hay una mayor cantidad de consultas distribuidas en una mayor cantidad de dominios.

7.2.7 Reglas de asociación por dominio

Se continúa el análisis de los dominios aplicando técnicas de minería de datos, se utiliza el software Rapid Miner para determinar reglas de asociación de dominios consultados en los tres años. Se utiliza el modelo de asociación y conjuntos frecuentes, específicamente los operadores FP-growth y Create Association Rules. Se consideró un soporte mínimo de 0.15.

La idea es identificar como consultan nuestros usuarios los dominios, si existe alguna asociación entre dominios consultados, cuando consultan más de un dominio.

Se generaron tres procesos en los que se obtuvo las reglas de asociación de los dominios consultados para cada año, 2012 a 2014. Para comparar estas reglas de asociación, se seleccionan las reglas con más alto soporte (frecuencia) y confianza (certeza de que aparezcan juntas la premisa y la conclusión), obtenidas para cada año, identificando la premisa y la conclusión asociada, el soporte y la confianza. En la siguiente tabla se identifica un extracto con las 10 primeras reglas para cada año. En el Anexo 4 se puede consultar un listado de hasta 100 reglas de asociación generadas para cada año.

Tabla 7-28: Reglas de asociación por dominios consultados 2012

Premisa	Conclusión	Soporte	Confianza
content.ebscohost.com	web.ebscohost.com	0,62676282	1
media.proquest.com	search.proquest.com	0,46169872	1
www.sciencedirect.com, content.ebscohost.com	web.ebscohost.com	0,43910256	1
content.ebscohost.com, search.proquest.com	web.ebscohost.com	0,43862179	1
web.ebscohost.com, media.proquest.com	search.proquest.com	0,40192308	1
content.ebscohost.com, media.proquest.com	web.ebscohost.com	0,35785256	1
content.ebscohost.com, media.proquest.com	search.proquest.com	0,35785256	1
content.ebscohost.com, media.proquest.com	web.ebscohost.com, search.proquest.c	0,35785256	1
web.ebscohost.com, content.ebscohost.com, media.proquest.com	search.proquest.com	0,35785256	1
content.ebscohost.com, search.proquest.com, media.proquest.com	web.ebscohost.com	0,35785256	1

La regla con más alto soporte y confianza para el año 2012 es la que nos indica que cada vez que se consulta *content.ebscohost.com* se consulta también *web.ebocohost.com*, lo que puede indicar una relación de subordinación de ambos sitios ya que pertenecen a un dominio similar. Lo mismo sucede con la segunda regla, cada vez que se consulta *media.proquest.com* se consulta también *search.proquest.com*. En la tercera regla recién se observa una relación de sitos con dominios distintos, al relacionar *www.sciencedirect.com* y *content.ebscohost.com* se consulta también *web.ebocohost.com*, pero se mantiene la relación del domino Ebscohost. En las 10 primeras reglas derivadas para 2012 se relacionan sitios de dominios asociados, los que pueden ser páginas subordinadas a un mismo dominio con distintos nombres.

Tabla 7-29: Reglas de asociación por dominios consultados 2013

Premisa	Conclusión	Soporte	Confianza
media.proquest.com	search.proquest.com	0,43416493	1
search.proquest.com, content.ebscohost.com	web.ebscohost.com	0,37318087	1
web.ebscohost.com, media.proquest.com	search.proquest.com	0,36573112	1
www.sciencedirect.com, content.ebscohost.com	web.ebscohost.com	0,36157311	1
www.sciencedirect.com, media.proquest.com	search.proquest.com	0,32241857	1
content.ebscohost.com, media.proquest.com	web.ebscohost.com	0,30266805	1
content.ebscohost.com, media.proquest.com	search.proquest.com	0,30266805	1
content.ebscohost.com, media.proquest.com	web.ebscohost.com, search.proquest.	0,30266805	1
web.ebscohost.com, content.ebscohost.com, media.proquest.com	search.proquest.com	0,30266805	1
search.proquest.com, content.ebscohost.com, media.proquest.com	web.ebscohost.com	0,30266805	1

Para el año 2013 la regla con mayor soporte y confianza es la que indica que cada vez que se consulta *media.proquest.com* se consulta también *search.proquest.com*, siguiendo el patrón del año anterior, se observa que ambos son sitios con un dominio común pero con distinto nombre, lo que puede indicar una subordinación de los sitios consultados. Lo mismo sucede con el resto de las primeras reglas para el año. En la premisa se asocian sitios de

dominios distintos pero en la conclusión siempre hay un sitio con un dominio similar a alguno existente en la premisa.

Tabla 7-30: Reglas de asociación por dominios consultados 2014

Premisa	Conclusión	Soporte	Confianza
ebscohost.com, els-cdn.com	sciencedirect.com	0,28775316	1
els-cdn.com, proquest.com	sciencedirect.com	0,26097594	1
els-cdn.com, springer.com	sciencedirect.com	0,22019845	1
ebscohost.com, els-cdn.com, proquest.com	sciencedirect.com	0,2123148	1
ebscohost.com, ac.els-cdn.com	www.sciencedirect.com	0,19206198	1
els-cdn.com, wiley.com	sciencedirect.com	0,18608128	1
pubmed.gov	nih.gov	0,18200353	1
proquest.com, ac.els-cdn.com	www.sciencedirect.com	0,17371211	1
ac.els-cdn.com, link.springer.com	www.sciencedirect.com	0,17208101	1
ac.els-cdn.com, content.ebscohost.com	www.sciencedirect.com	0,16868289	1

Para el año 2014 se rompe el patrón de asociar sitios de dominios similares, la primera regla, con un soporte mucho menor que las 10 primeras reglas del 2012 y 2013, nos indica que cada vez que se consulta *ebscohost.com* y *els-cdn.com* se consulta también *sciencedirect.com*. Sitios con distintos dominios en la premisa y la conclusión. El resto de las 10 primeras reglas de asociación para este año muestran asociación en la consulta de los distintos dominios.

Este patrón se podría explicar pesando en la incorporación de nuevas bases de datos el año 2014, lo que aumentó y diversificó la cantidad de consultas.

7.2.8 Clustering de consultas y notas

Luego se aplican algoritmos de clustering, considerando las variables notas, cantidad de sesiones y duración de estas.

Se representan los resultados en las matrices de dispersión obtenidas para cada algoritmo aplicado por año, considerando las tres variables en estudio y sus combinaciones que incluyen notas y sesiones, notas y duración, sesiones y notas, sesiones y duración, duración y notas, duración y sesiones.

Finalmente, se genera un set de datos para cada año con las notas (en Zscore por carrera), cantidad de sesiones y duración promedio por alumno, para el 2012 se generó un archivo con 18206 alumnos, para el 2013 fueron 21343, y para el 2014 fueron 21874.

Estos archivos se procesan en Rapid Miner aplicando en primer lugar el algoritmo de clustering K-means, luego de probar varios parámetros se deja un parámetro k=4. Luego se aplica a los mismos conjuntos de datos el algoritmo DBSCAN, considerando los parámetros Epsilon=20 y min point=10.

7.2.7.1 K-means (k=4)

Al aplicar el algoritmo a los datos del 2012 se obtuvo clusters distribuidos de la siguiente forma.

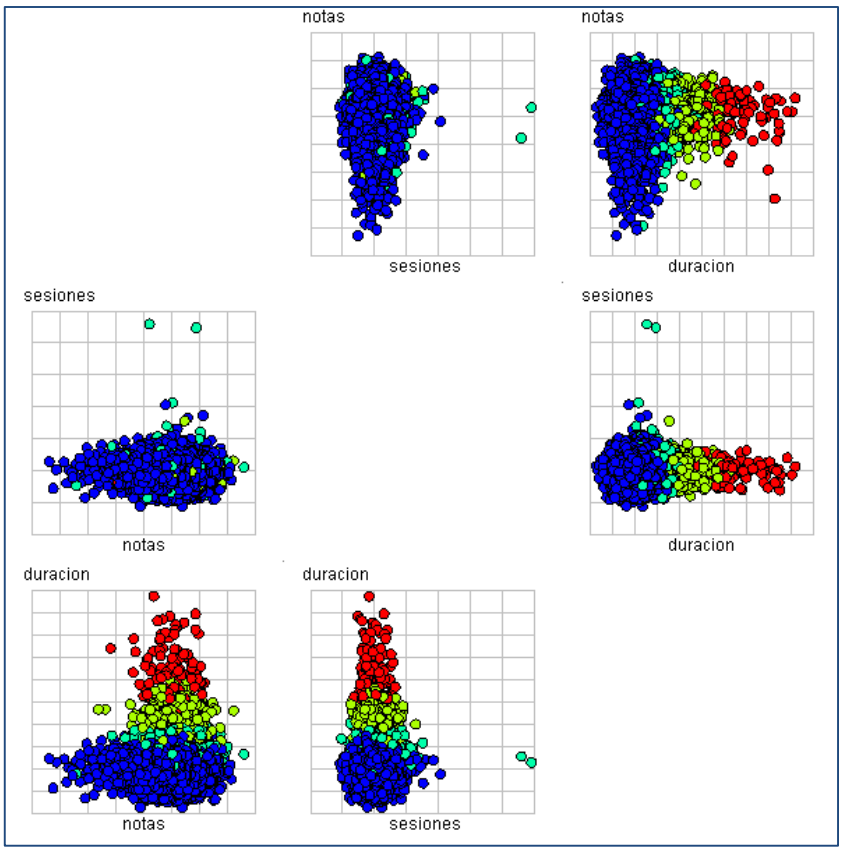

Imagen 7—47: Clusters de notas, sesiones y duración con K-means para 2012

En la relación notas-sesiones y sesiones-notas, no se aprecian en forma distribuida los clusters. Se observa algunos puntos dispersos, pero la mayoría se concentra en el clúster azul, un grupo con baja cantidad de sesiones y con una distribución de notas en todo el rango, tiende a ser más denso en las notas superiores.

En esta distribución se observa un leve patrón que indica que quienes tienen mayor cantidad de sesiones tienden a tener mejores notas.

En la relación notas-duración de sesiones y duración-notas, se observan claramente al menos tres de los clusters, el primer clúster, en azul, con la mayor densidad corresponde a los alumnos con una duración de sus sesiones menor a la media y una distribución de notas con una leve tendencia a estar sobre la media. El segundo clúster, en calipso, es el de menor tamaño, se aprecia una duración de sesiones bajo la media con notas alrededor de la media y superiores. El tercer clúster, en amarillo, presenta una mayor densidad con una duración de sesiones alrededor de la media y notas mayoritariamente sobre la media. El

cuarto clúster, en rojo, con una densidad menor que el tercero, presenta la mayor duración de las sesiones, por sobre la media, y notas también sobre la media.
Se aprecia en esta distribución que los mejores alumnos son quienes presentan sesiones de mayor duración, no existen alumnos de notas bajo el promedio que tengan una duración sobre la media.

Respecto a la relación duración-sesiones, se observa los cuatro clúster bien distribuidos, el primer clúster, en azul, de mayor densidad, incluye a quienes tienen una cantidad de sesiones y una duración de estas menores a la media. El segundo clúster, en calipso, menor en densidad, presenta una cantidad de sesiones menor a la media, una duración levemente bajo la media, incluye algunos casos extremos (layouts) con una alta cantidad de sesiones. El tercer clúster, en amarillo, presenta a quienes tienen una duración de sesiones alrededor de la media y la cantidad de consultas, como el resto, también bajo la media. El cuarto clúster, en rojo, incluye a quienes tienen una alta duración de las consultas y una cantidad de sesiones bajo la media.

Se aprecia que en esta distribución la mayoría de los usuarios se encuentra bajo la media en la cantidad de sesiones.

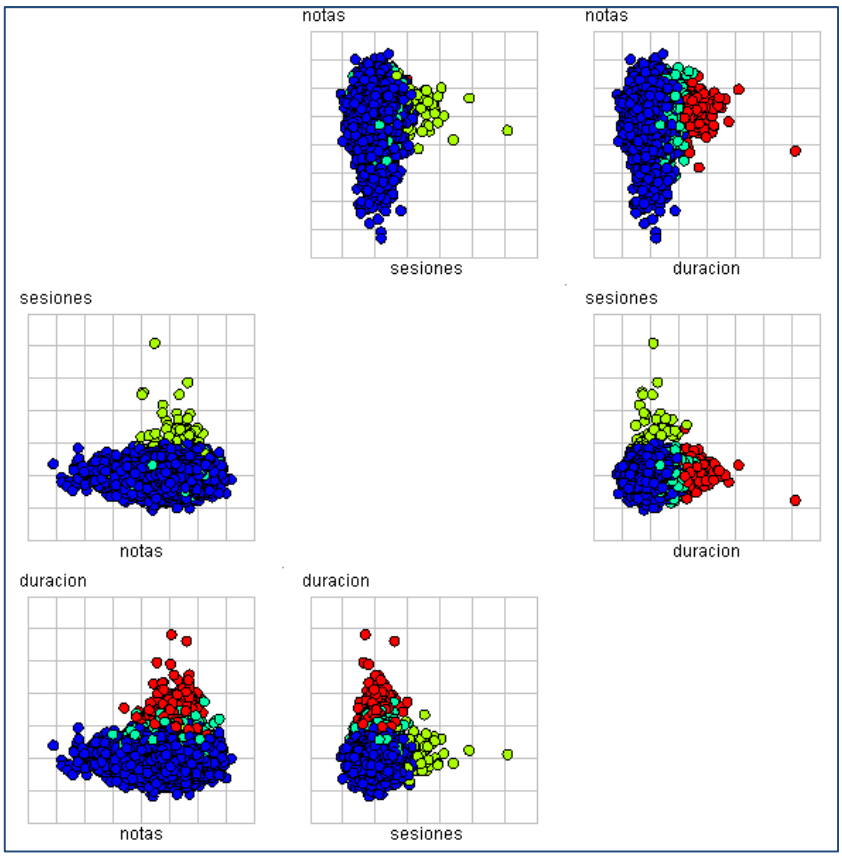

Imagen 7—48: Clusters de notas, sesiones y duración con K-means para 2013

Para el año 2013 en la relación notas-sesiones y sesiones-notas, se observan tres de los cuatro clusters, el primer clúster, en azul, es el que presenta mayor densidad, se observa una tendencia a tener pocas sesiones en todo el rango de notas, aunque al igual que en 2012, se aprecia una leve tendencia a un aumento en las notas en la medida que aumenta la cantidad de sesiones. El clúster en calipso, se aprecia brevemente en un par de puntos, que sólo indican notas sobre el promedio moviéndose en la media de la cantidad de consultas. El clúster en amarillo, indica claramente un grupo menor de usuarios que con notas por sobre la media tiene una cantidad de sesiones alta, no son las notas más altas pero presentan el patrón de que están por sobre el promedio y son quienes más consultan. El clúster rojo sólo es posible apreciarlo en un punto, ya que su densidad es muy baja, con alta cantidad de sesiones y alto promedio de notas.

En esta distribución se observa una tendencia, los usuarios que presentan una cantidad de consultas por sobre el promedio tienen mejores notas.

En la relación notas-duración y duración-notas, se observa sólo tres de los cuatro clusters, en el primero, azul, un grupo con alta densidad, presenta una duración de las sesiones por debajo del promedio con notas distribuidas a lo largo de todo el rango, con una leve tendencia hacia la parte superior de la media. En el segundo clúster, calipso, de menor densidad se observa que las notas van desde el promedio hacia arriba y la duración corresponde a los alrededores del promedio. En el tercer clúster, rojo, se encuentran quienes presentan una duración de las sesiones superior al promedio y tienen notas por sobre el promedio.

En esta distribución se aprecia una tendencia clara, aquellos que tienen las sesiones con mayor duración son quienes presentan mejores notas.

Respecto a la relación duración-sesiones y sesiones-duración, se observan claramente los cuatro clúster, en el primero, azul, el más denso, se encuentran quienes tienen pocas sesiones y de baja duración, En el clúster calipso, se encuentran quienes tienen pocas sesiones con una duración mayor que el clúster azul. En el clúster amarillo, se encuentran quienes tienen mayor cantidad de sesiones con una duración por debajo de la moda. En el clúster 4, rojo, se encuentra el grupo con mayor duración de las sesiones, sobre el promedio, pero con pocas sesiones (bajo el promedio). En esta distribución se pudo observar que la tendencia es a tener menor cantidad de sesiones de corta duración, y en los extremos se observa una asociación inversa cuando existen altas sesiones son de corta duración y hay pocas sesiones de alta duración.

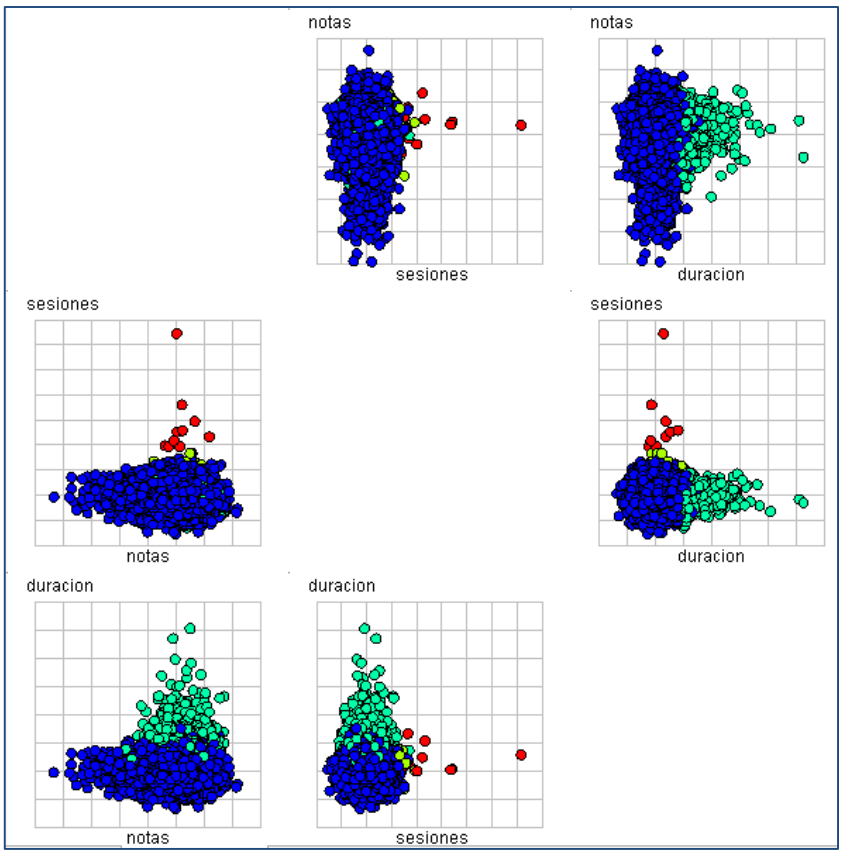

Imagen 7—49: Clusters de notas, sesiones y duración con K-means para 2014

En la relación notas-sesiones y sesiones-notas, se observan para el año 2014 los cuatro clúster, pero el primer clúster, en azul, es el que presenta la mayor densidad, se observa una tendencia a tener pocas sesiones en todo el rango de notas, aunque al igual que en 2012 y 2013, se aprecia una tendencia a un leve aumento en las notas en la medida que aumenta la cantidad de sesiones, también se aprecia una mayor densidad en el sector que corresponde a quienes tienen mejores notas, por sobre la media. En el clúster en calipso, se aprecia solo en dos puntos, de los que se puede decir que corresponden a notas altas con pocas consultas. En el clúster amarillo, se aprecia en un par de puntos que indican buenas notas con mayor cantidad de consultas. El clúster rojo, su densidad es muy baja respecto al clúster azul, pero indica claramente un grupo con altas notas y alto nivel de consultas.

En esta distribución se observa la misma tendencia del 2013, los usuarios que presentan una cantidad de consultas alta tienen mejores notas.

En la relación notas-duración y duración-notas, se observa sólo dos de los cuatro clusters, en el primero, azul, un grupo con alta densidad, presenta una duración de las sesiones por debajo del promedio con notas distribuidas en todo el rango, con una leve tendencia a aumentar la duración al aumentar la nota, ya que se observa mayor densidad en el sector de las notas más altas. En el segundo clúster, calipso, de menor densidad y mayor distribución, se observa, al igual que en 2012 y 2013, que quienes tienen sesiones con mayor duración, sobre el promedio, son también quienes tienen las notas más altas, desde el promedio hacia arriba.

En esta distribución se aprecia una tendencia clara, los alumnos con mejores notas consultan en sesiones de mayor duración. No se observan alumnos de notas bajas que consulten en sesiones de alta duración.

Respecto a la relación duración-sesiones y sesiones-duración, se observan los cuatro clusters, el primero, azul, más denso y con una distribución aglomerada, incluye a quienes tienen pocas sesiones y de baja duración. En el clúster calipso, que es de menor densidad y más distribuido que el primer clúster, se encuentran quienes tienen pocas sesiones y de alta duración. El clúster amarillo, es muy pequeño, un par de puntos aglomerados, incluye a quienes tienen mayor cantidad de sesiones con una duración menor, por debajo del promedio. En el clúster rojo, se encuentra el grupo con la mayor cantidad de sesiones, pero de corta duración, bordeando el promedio.

En esta distribución se observa que se mantiene la tendencia a tener menor cantidad de sesiones pero de mayor duración.

En la tabla a continuación se puede observar las distancias entre los clusters, lo que indica la densidad de los clusters y la cantidad de puntos que cubren. Se indica los colores de su visualización en las figuras anteriores.

Tabla 7-31: Vector de desempeño de los clusters con K-means

Vector de desempeño de los clusters	2012	2013	2014
Promedio de distancia interclúster:	-359.876.114	-1.270.109.009	-3.633.416.052
Distancia promedio en clúster 1:	-377.530.084	-541.013.814	-1.408.456.216
Distancia promedio en clúster 2:	-56.753.402	-125.472.699	-457.575.264
Distancia promedio en clúster 3:	-140.171.515	-1.330.333.997	-129.070.983
Distancia promedio en clúster 4:	-226.918.708	-224.992.808	-3.796.385.850

El año 2012 es el que presenta las distancias intercluster más pequeñas, es decir es que presenta más aglomeración de los puntos o menos puntos en estudio, ya que la cantidad de usuarios que consultaron ese año fue menor que los otros dos años.

El año 2014 es para los cuatro clúster el que tiene las distancias más grandes, al tener mayor cantidad de alumnos con sesiones.

7.2.7.2 DBSCAN

Los parámetros utilizados son un Epsilon de 20 y un punto mínimo de 10.

En general este algoritmo presentó un desempeño menos eficiente en la generación de grupos, lo que podría explicarse porque la distribución de los puntos es más bien homogénea, con grandes focos de mayor densidad y poca distribución.

En la matriz de dispersión se observa claramente la densidad de los puntos y la distribución de los clusters generados.

Para el año 2012 el algoritmo generó ocho clusters, que se pueden apreciar sólo en algunas de las relaciones entre las variables.

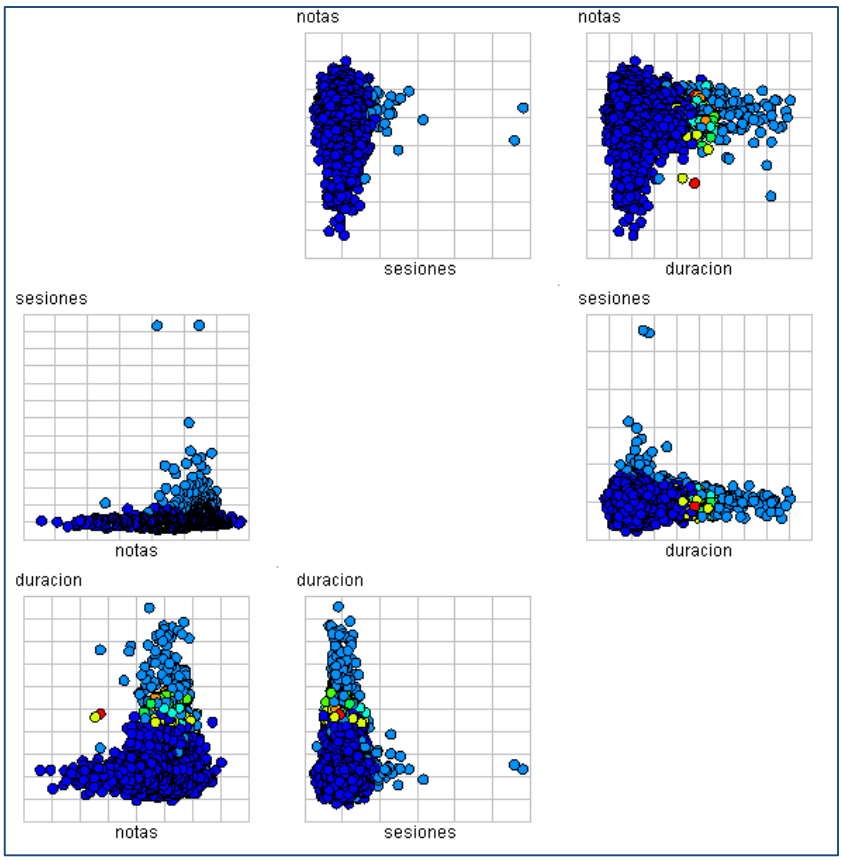

Imagen 7—50: Clusters de notas, sesiones y duración con DBSCAN para 2012

Para la relación notas-sesiones y sesiones-notas, se aprecia un gran clúster, en azul, con la distribución más densa de los puntos, que corresponde a quienes han realizado pocas sesiones y sus notas se distribuyen a lo largo del rango. El segundo clúster, que se puede ver en celeste, son los puntos con una distribución mayor y baja densidad, que corresponden a quienes poseen una mayor cantidad de sesiones y tienen notas por sobre el promedio. En esta distribución se observa la tendencia de que la mayor cantidad de sesiones la generan quienes tienen las notas por sobre el promedio.

Para la relación notas-duración y duración-notas, se aprecian ocho clusters, de los cuales el mayor, en azul, agrupa los puntos más densos, considera a quienes tienen sesiones d poca duración y han obtenido notas en todo el rango, se observa si una leve tendencia tener notas sobre el promedio. El segundo en densidad corresponde al celeste, que corresponde a quienes han realizado sesiones de larga duración y presentan notas sobre el promedio.

El resto de los clusters generados se distribuyen entre quienes han presentado un tiempo promedio de duración y sobre el promedio de las notas, con dos puntos bajo el promedio. En esta distribución es evidente la tendencia de que quienes tienen las sesiones de duración más alta son quienes tienen, en general, mejores notas.

Para la relación sesiones-duración y duración-sesiones, se observan los ocho clusters, de los cuales dos son los mayores, el azul, grupo mayoritario, agrupa a quienes tienen pocas sesiones y de corta duración; el celeste agrupa dos tipos de alumnos, por un lado a quienes tienen pocas sesiones de alta duración, y a quienes tienen mayor cantidad de sesiones pero de corta duración. El resto de los clusters se distribuye entre quienes se encuentran en el promedio de la duración de sesiones con una corta duración.

En esta distribución se observa que la mayoría de los alumnos realizan pocas sesiones, pero su duración varía, aunque se observa una tendencia a que son de corta duración.

Para el año 2013 el algoritmo generó 38 clusters para este año, de los cuales sólo los mayores se pueden apreciar en la matriz de dispersión.

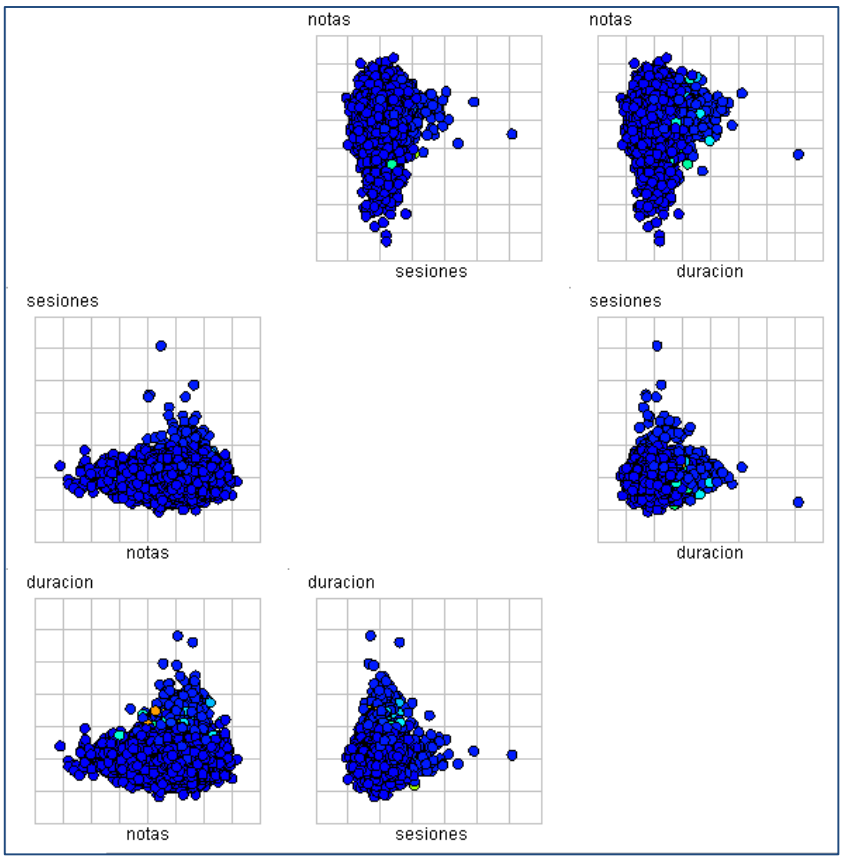

Imagen 7—51: Clusters de notas, sesiones y duración con DBSCAN para 2013

Para la relación notas-sesiones y sesiones-notas, se observa un gran clúster azul que comprende la mayoría de los puntos, y algunos menores, sólo dos puntos visibles en la matriz, que se ubican en la media de las notas. En esta distribución se observa que los alumnos que más sesiones tienen son aquellos con notas sobre el promedio, no se observan alumnos con alta cantidad de sesiones y bajas notas.

Para la relación notas-duración y duración-notas, se aprecia un gran clúster azul, que agrupa la mayoría de los puntos, y algunos pocos puntos de distintos clusters que se ubican en la media de la duración y distribuidos en las notas, con una tendencia por sobre el promedio. En esta distribución también se aprecia que los alumnos con sesiones de duración más altas son aquellos que tienen notas sobre el promedio, no se observan alumnos de notas bajas y sesiones de alta duración.

Para la relación duración-sesiones y sesiones-duración, se observa un gran clúster azul en la mayor parte de la distribución de puntos, con algunos pocos clusters distribuidos por

sobre la media de la duración y bajo la media de las sesiones. En esta distribución se observa un gran foco pocas sesiones de baja duración, pero en los extremos es posible observar una asociación inversa cuando existe alta cantidad de sesiones hay baja duración y cuando existe una alta duración hay baja cantidad de sesiones.

Para el año 2014 el algoritmo generó 22 clusters, siendo el azul el principal y los otros más pequeños.

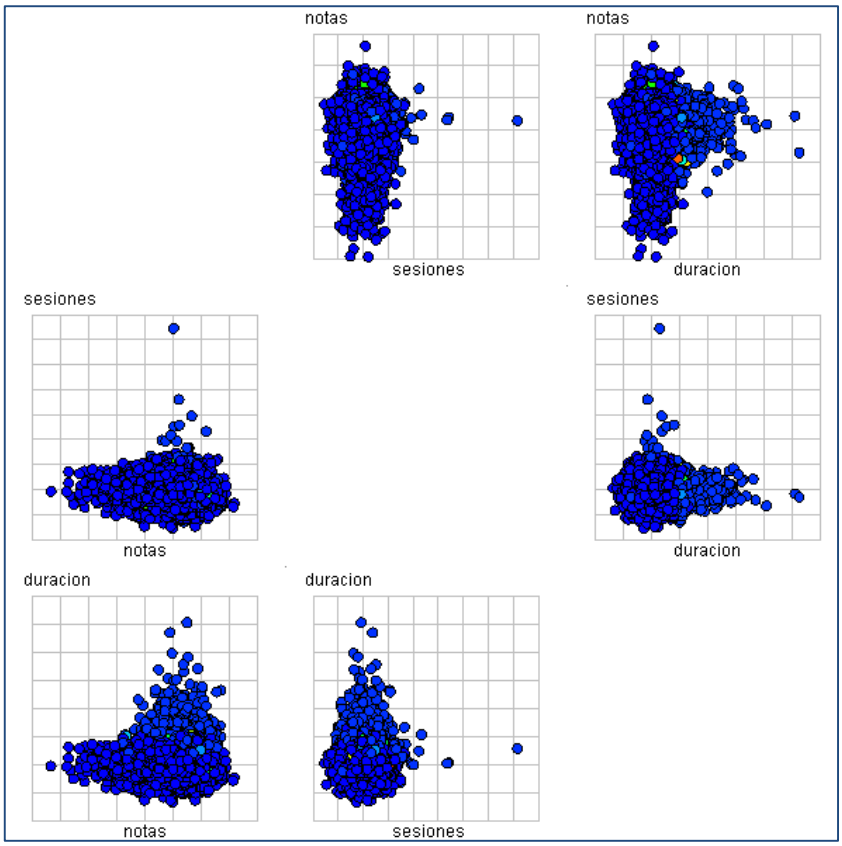

Imagen 7—52: Clusters de notas, sesiones y duración con DBSCAN para 2014

Para la relación notas-sesiones y sesiones-notas, se observa principalmente el clúster azul, con la excepción de un punto de otro clúster ubicado en altas notas y pocas sesiones. La distribución se centra en un grupo a lo largo del rango de notas con baja cantidad de sesiones, existen si algunos casos excepcionales de alta cantidad de sesiones y notas sobre el promedio. No se observan casos de bajas notas y alta cantidad de sesiones.

Para la relación notas-duración y duración-notas se observa el gran clúster azul, y algunos puntos de otros clusters, unos en el promedio de notas y duración otro en notas altas con

baja duración. La distribución nos indica una fuerte tendencia a las sesiones de alta duración de alumnos con notas por sobre el promedio. No se observan alumnos de bajas notas con sesiones de alta duración.

Para la relación sesiones-duración y duración-sesiones, se observa el gran clúster azul en la mayoría de la distribución, con dos excepciones en el promedio de duración y bajas sesiones. La distribución indica una densidad mayor entre quienes tienen pocas sesiones y de corta duración. En los extremos se observan más casos de alta duración y pocas sesiones que alta cantidad de sesiones de corta duración.

Para observar la distancia entre los clusters más destacados se confecciona una tabla comparativa de los cinco primeros clusters obtenidos para cada año y las distancias promedio de cada uno de ellos en cada año.

Tabla 7-32: Vector de desempeño para clusters con Dbscan

Vector de desempeño para los clusters	**2012**	**2013**	**2014**
Distancia promedio interclúster	**-850.624.352**	**-1.994.504.767**	**-1.803.229.352**
Distancia promedio en clúster 1, 1, 0	**-886.675.146**	**-2.059.108.127**	**-4.931.555.035**
Distancia promedio en clúster 0, 0, 1	**-474.168.797**	**-1.278.959.467**	**-1.411.754.903**
Distancia promedio en clúster 3, 10, 7	**-6.553.072**	**-2.229.100**	**-4.071.711**
Distancia promedio en clúster 5, 5, 3	**-2.237.731**	**-1.660.383**	**-3.526.287**
Distancia promedio en clúster 10, 7, 5	**-563.034**	**-1.351.090**	**-1.938.523**

El primer clúster de cada año, en azul, es el que posee la distancia menor entre los puntos, es decir es el más denso. El resto de los clusters se encuentran con una distancia mucho mayor, y por esta razón menos visibles en la matriz de distribución.

8. Conclusiones

A continuación se presentarán los hallazgos, tendencias y patrones, con los que se pretende contestar las preguntas de investigación y cumplir con los objetivos de este estudio.

8.1 Colección presencial

Los materiales con mayor cantidad de consultas son del área de Humanidades, Cs. Naturales y Cs. Sociales, el material de Cs. Médicas e Ingeniería tiene menor demanda. A través de los cinco años de estudio, se observa una leve disminución de las transacciones, en todas las áreas.

Respecto a la cantidad de áreas consultadas, se observa una tendencia al alza de la proporción de solicitudes de material de 1 o 2 disciplinas respecto a la solitud de 3 o más disciplinas Las solicitudes de material de tres y más áreas distintas, presenta un leve descenso en el tiempo de estudio. Existe una relación negativa entre la cantidad de áreas que consultadas por un usuario y la cantidad de usuarios.

Las consultas de material de sólo un área corresponden a casi un 40% del total de las consultas, mientras las solicitudes de 2 y más áreas consideradas en su conjunto corresponden a un 60% del total. Se observa que las consultas en 1 y 2 áreas corresponden al 74,19% de las consultas realizadas por nuestros usuarios, siendo las consultas en 2 áreas levemente superior, y sólo un cuarto de las consultas corresponden a 3, 4 y 5 áreas.

Si se considera el área del usuario, de acuerdo a su unidad Académica, y se compara con las consultas de material en su área y las consultas en otras áreas se observa que a lo más un 38% de los usuarios consultan más de otras áreas que la propia, y al menos un 61% consultan más de su área que de otras que no corresponden a su Unidad Académica. Se observa que las Unidades académicas que consultan más de otras áreas que de la propia, corresponden a aquellas que tienen un plan común de ramos de otras unidades académicas, como por ej. Cálculo, Física, Química.

Respecto a las consultas por áreas del material, las áreas de Cs. Sociales y Humanidades las que se consultan mayoritariamente en forma exclusiva y combinada, seguidas por la combinación con el área de Cs. Naturales.

Al analizar conjuntos frecuentes de las áreas consultadas, se repite este patrón, siendo las áreas menos consultadas las de Ingeniería y Cs. Médicas, tanto en forma exclusiva como combinadas. Hasta el año 2011 se observa que el área con más consultas era Cs. Sociales y luego Humanidades, a partir del 2012 esta secuencia se invierte.

Al considerar la entropía de las consultas de estas áreas en los cinco años de estudio, nos encontramos con que el área que presenta un patrón más homogéneo es Cs. Sociales, al considerar dos áreas consultadas es la dupla Cs. Sociales-Humanidades, siendo las que se mantienen con la mayor cantidad de consultas a través de los años. Si estas se asocian a las otras áreas con baja entropía, suben la entropía del conjunto.

En las reglas de asociación obtenidas se puede observar, con los niveles más altos de confianza del conjunto, que cada vez que se consulta material de dos áreas, y una de ellas es Cs. Sociales, la otra será Humanidades, y viceversa. Luego se suma la combinación de cualquiera de las dos áreas anteriores con Cs. Naturales. Este patrón se observa con frecuencia en los cinco años de estudio, con ciertas variaciones en el orden, pero considerando las tres áreas mencionadas.

Al consultar tres áreas, la regla con más alta confianza que se detectó, mucho menor que las reglas anteriores, es la que indica que cuando se pide material de Cs. Naturales se pide también material de Cs. Sociales y Humanidades, y viceversa.

Se descubre entonces una tendencia clara a subir las consultas en una disciplina o dos, y una tendencia a disminuir las consultas de más de tres áreas.

Al considerar el área del usuario se detecta que carreras como Cs. Naturales, Cs. Médicas y Humanidades, piden mayoritariamente material de su área. Los de Ciencia y Tecnología piden material de otra área más que de su propia área. Los usuarios de Humanidades y Cs. Naturales son quienes tienen el promedio más alto de consultas, el resto de las áreas se mantienen por debajo del promedio de la Universidad. Los usuarios de Cs. Sociales son

quienes piden más material de 1, 2 y 3 áreas, los usuarios de Ciencia y tecnología son quienes tienen mayor cantidad de consultas en 3 y 4 disciplinas distintas.

En relación a la asociación entre notas obtenidas por los alumnos y la multidisciplinariedad de sus consultas (cantidad de áreas consultadas), se observa una tendencia a aumentar el p-value y disminuir el chi cuadrado en los años estudiados. La asociación sólo es muy leve para el año 2012, luego no se observa una diferencia significativa y no existe dependencia entre las variables.

A nivel más específico, considerando las carreras de los usuarios, se observa que en todas las carreras se solicita material de distintas disciplinas, no existen carreras que pidan exclusivamente de su área, aunque se observa una tendencia a consultar mayoritariamente de un área o dos. Existe una tendencia en las carreras humanistas, seguidas por las de Ciencias exactas, a tener una mayor cantidad de transacciones en biblioteca, por otro lado las carreras de Ingeniería y Cs. de la salud son las con menos consultas en biblioteca.

En todas las áreas de los usuarios se aprecia una leve disminución de las transacciones a través de los años. La baja se presenta fuertemente en las áreas con mayor cantidad de transacciones y más leve en las áreas que piden menos. A través de los años se refleja en los gráficos la disminución de aquellos usuarios con altas transacciones.

Se observa que en los tres años los usuarios con notas por sobre el promedio que consultan en biblioteca corresponden al 54%, de estos alrededor del 16,5% corresponden al uso de biblioteca por sobre el promedio. Por el otro lado, cerca del 32% de los alumnos usan la biblioteca sobre el promedio, de estos quienes tienen un zscore sobre el promedio corresponden a un 18%, más que aquellos con notas bajo el promedio.

El p-value para los tres años es< 0.05, por lo que la diferencia se considera estadísticamente significativa.

Al aplicar algoritmos de clustering relacionando transacciones y notas de los alumnos, se encontraron agrupaciones de usuarios con las siguientes características:

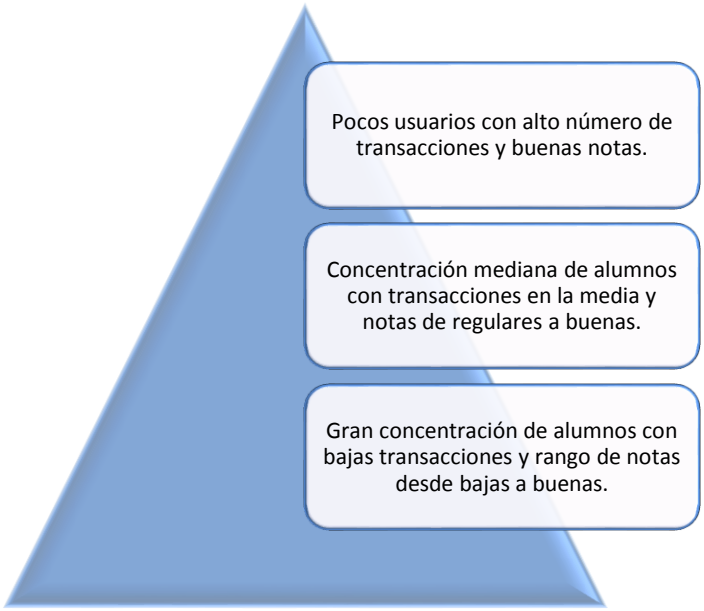

Imagen 8—1: Grupos obtenidos en los clusters de transacciones presenciales y notas

No se observan grupos de alumnos con alta cantidad transacciones y que a la vez tengan bajas notas, existe una tendencia clara a que en la medida que sube la cantidad de transacciones los alumnos presentan notas altas. También se observa que quienes tienen menos transacciones se distribuyen en todo el rango de notas.

8.2 Consultas a la colección electrónica

Desde el año 2012 al 2014 se observa un aumento de 2,56 veces (256%) de usuarios que consultan la colección electrónica. La cantidad de consultas también presenta una fuerte tendencia a aumentar, ya que se ha elevado 21,6 veces desde el 2012 al 2014. Si se considera el promedio de consultas por usuario al año, este ha subido 8 veces (847%). El tiempo de duración de las consultas también ha ido aumentando 4,6 veces (465%) estos años.

Se puede deducir entonces que una vez que el usuario comienza a usar la colección electrónica, hay una fuerte tendencia a ir aumentando la cantidad de sesiones que realiza y con mayor tiempo de duración.

Los usuarios del área de Cs. Médicas son quienes más consultas realizan seguidos por los de Cs. Naturales, tendencia que se mantiene en los años. Todas las áreas han crecido 10 veces aproximadamente (un 1000%) en promedio de la cantidad de sesiones, excepto Cs, Médicas cuyas consultas promedio han aumentado 5,4 veces (543%).

En cuanto a la duración de las sesiones por área, es Cs. Sociales la que tiene el mayor aumento desde el 2012 al 2014, con una subida de 9,24 veces (924%), Cs. Médicas tiene el menor aumento con 1,56 veces (156%), aunque igualmente sigue siendo el área con las sesiones de mayor duración para los tres años.

Al considerar el promedio por usuario, son los usuarios de Humanidades quienes presentan pocas sesiones al año y la mayor duración por sesión, las sesiones más cortas las tienen los usuarios de Cs. Naturales.

Si tomamos en cuenta las carreras de los usuarios, en general las carreras de Medicina y Derecho se encuentran entre las que presentan la mayor cantidad de sesiones y duración. Al comparar las sesiones de la colección electrónica y las consultas en biblioteca a través de los años de carrera nos encontramos con que el uso de la biblioteca presencial es alto los primeros años de carrera y luego desciende levemente. Las sesiones de la colección electrónica luego de una baja inicial, tienden a subir con los años de carrera, especialmente a partir del tercer año, al igual que la duración de las sesiones, y bajan a partir del séptimo

año. Esta tendencia es observada en los tres años (2012-2014), con algunas variaciones. Mientras el uso de la biblioteca presenta un levo descenso en los años el uso de la colección electrónica presenta un abrupto aumento.

El promedio de consultas ha ido aumentando rápidamente al igual que la cantidad de usuarios que consulta la colección electrónica.

En todos los años se observa la tendencia de que aunque sólo el 22% aproximadamente de los usuarios consultan sobre el promedio, la mayoría de ellos (14%), tienen notas por sobre el promedio.

El p-value obtenido es $2.2x10^{-16}$, mucho menor a 0.05, lo que indica que existe una diferencia estadísticamente significativa, y por lo tanto se rechaza la hipótesis nula que establece que no existe diferencia entre el rendimiento académico de los alumnos que usan los servicios de la biblioteca por sobre el promedio. .

Al comparar la cantidad de consultas en la biblioteca y las sesiones en la colección electrónica, se observa que la tendencia general es a que quienes consultan poco la biblioteca consulta poco también la colección electrónica. Sólo un 2% de los usuarios consultan sobre el promedio la colección electrónica y la biblioteca. El p-value de estas dos variables para los años 2012-2014 fue <0.05, por lo que existe una diferencia significativa entre ellas.

En la medida que aumentó la cantidad de consulta electrónicas aumentó también la cantidad de dominios consultados el 2013 y luego bajó levemente el año 2014, lo que va en relación con el aumento de recursos incorporados a la colección y a que los recursos ya probados reciben mayor cantidad de consultas. Existen entre diez a quince dominios que reciben la mayor cantidad de consultas, los que a pesar de haber presentado variaciones de nombres o derivado nuevos sitios presentan un patrón de uso alto durante el tiempo de estudio. Entre ellos Ebscohost, Sciencedirect, Proquest y Springer. Existe una mayor cantidad de consultas distribuidas en una mayor cantidad de dominios.

Las reglas de asociación nos muestran en primer lugar dominios derivados o asociados los dos primeros años, y ya el 2014 las primeras reglas, con soporte menor a la de los años

anteriores, corresponden a dominios distintos (Ebscohost-Proquest, Els-Proquest-Sciencedirect). En general se observa una asociación entre los recursos de distintos poveedores consultados por los usuarios, especialmente provenientes de Ebscohost, Sciencedirect, Proquest y Springer.

Al generar agrupaciones con datos de notas, sesiones y duración se encontraron agrupaciones de usuarios con diversas características.

Para la relación notas-cantidad de sesiones:

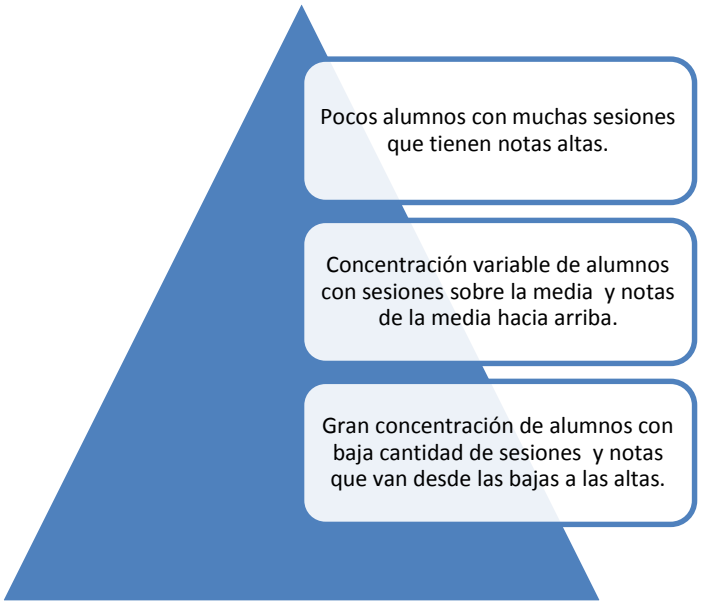

Imagen 8—2: Grupos obtenidos en los clusters de sesiones y notas

Se observa una leve tendencia a mayor cantidad de transacciones en usuarios con notas más altas. No se presentan casos de usuarios con bajas notas y alta cantidad de sesiones.

Para la relación notas-duración de las sesiones:

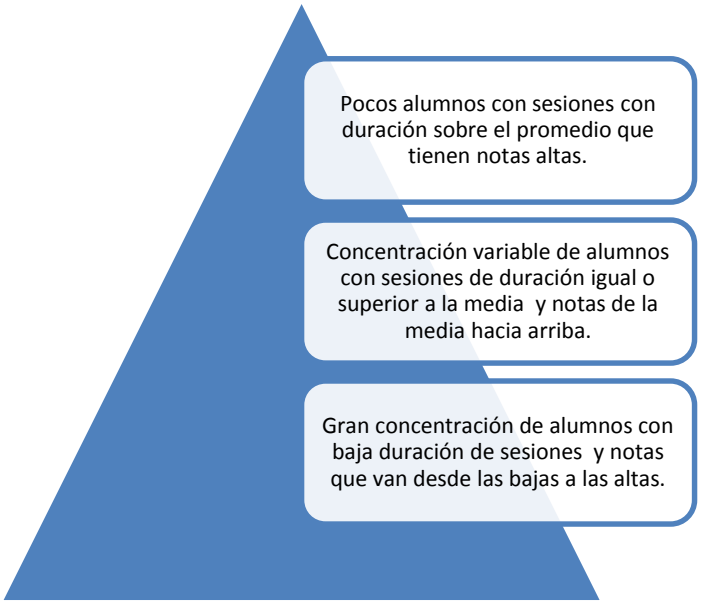

Imagen 8—3: Grupos obtenidos en los clusters de duración de sesiones y notas

Se observa una clara tendencia que los usuarios con mejores notas realizan sesiones de mayor duración. No se observan alumnos de bajas notas con sesiones de duración larga. Para la relación cantidad de sesiones-duración de las sesiones:

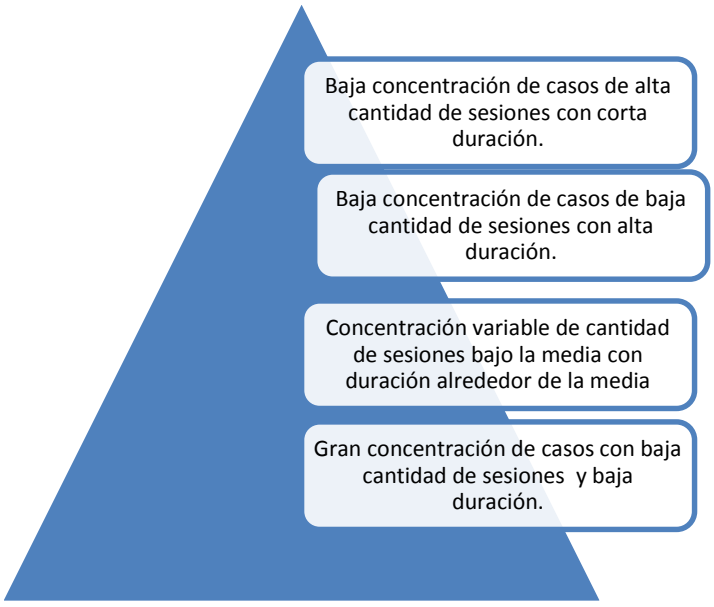

Imagen 8—4: Grupos obtenidos en los clusters de sesiones y duración

Se observa la tendencia a tener pocas sesiones y cortas. No hay casos de alta cantidad de sesiones con larga duración.

8.3 Contrastación de resultados con hipótesis

Al contrastar los resultados con las hipótesis planteadas al inicio del estudio, en general se validan las hipótesis, a continuación se verá en detalle cada una de ellas.

Hipótesis 1:

Ho: El 50% de las consultas son multidisciplinarias

H_1: Menos del 50% de las consultas son multidisciplinarias

Se observó que los usuarios, agrupados por el área de sus carreras, consultan material de su área en primer o segundo lugar, por lo general la segunda área consultada está asociada a la carrera, existe una baja consulta a otras áreas no relacionadas.

El 39% de los usuarios consultan material principalmente de una disciplina, las consultas de 2 o más disciplinas corresponden al 61%, de las cuales el 35% corresponde a consultas en 2 áreas.

A nivel de carrera se pudo observar que en ninguna carrera los usuarios consultan exclusivamente material de su área, al menos consultan otra área cercana. La multidisciplinariedad de las consultas no tiene una asociación con el rendimiento de los alumnos, en dos de los tres años estudiados, sólo para 2012 resultó con un p-value levemente significativo.

Se comparó el porcentaje de consultas multidisciplinarias respecto a las monodisciplinarias, por cada Unidad Académica. Durante los años 2010 y 2014, entre un 68% y un 61% de las unidades académicas tenía una mayoría de consultas monodisciplinarias, en promedio consultaban un 57% de material de sus áreas y un 43% de otras áreas. En el siguiente gráfico podemos observar que el porcentaje de las consultas multidisciplinarias va aumentando con los años, mientras que las consultas monodisciplinarias van disminuyendo.

Tabla 8-1: Análisis estadístico de la multidisciplinariedad de las consultas

Multidisciplinariedad	2010	2011	2012	2013	2014
>50%	**32,26%**	**32,26%**	**35,48%**	**35,48%**	**38,71%**
<50%	**67,74%**	**67,74%**	**64,52%**	**64,52%**	**61,29%**
Media	40%	41%	46%	43%	43%
Error típico	4%	5%	5%	4%	4%
Mediana	39%	37%	45%	43%	42%
Desviación estándar	25%	26%	27%	24%	24%
Mínimo	4%	4%	5%	4%	4%
Máximo	84%	87%	91%	88%	87%
p-value (promedio)	**2.2e-16**	**2.2e-16**	**2.2e-17**	**2.2e-16**	**2.2e-16**

El p-value de las consultas multidisciplinarias de cada unidad para todos los años estudiados fue siempre <0.05 lo que indica que existe una diferencia estadísticamente significativa, como para rechazar la hipótesis nula.

Hipótesis 2:

$H_{o:}$ No existe diferencia entre el rendimiento académico de los alumnos que usan los servicios de la biblioteca por sobre o bajo el promedio.

$H_{1:}$ Los alumnos con mayor rendimiento son los que usan los servicios de la biblioteca por sobre el promedio.

Respecto al uso de la biblioteca presencial, los usuarios con notas por sobre el promedio que usan la biblioteca corresponden a un 54% aproximadamente. Quienes usan la biblioteca sobre el promedio corresponden al 32%, de estos últimos, el 56% presentan notas sobre el promedio. Entre quienes usan la biblioteca bajo el promedio también se observa que la mayoría (53%) tiene notas sobre el promedio.

Tabla 8-2: Uso de biblioteca y notas 2012-2014

2012-2014 Uso de biblioteca y notas			
Promedios (sin uso cero)	Zscore <0,015	Zscore >=0,015	Total x bca
Uso bca >=22	44.04%	55.96%	24639
Uso bca<22	47.31%	52.69%	52179
Total x zscore	46.26%	53.95%	76818
X-squared = 62.663, df = 1, p-value = 2.453e-15			

El p-value para las notas de los alumnos y el uso de la biblioteca presencial de los tres años, 2012-2014, es <a 0.05, por lo que la diferencia se considera significativa, y se rechaza la hipótesis nula.

Respecto al uso de la colección electrónica, el promedio de consultas ha ido aumentando rápidamente al igual que la cantidad de usuarios que consulta la colección electrónica.
Cerca del 55% de quienes consultan tienen notas iguales o sobre el promedio anual del Zscore de notas. Aproximadamente el 78% de los alumnos que consultan lo hacen por debajo del promedio de consultas anuales. En todos los años se observa la tendencia de que aunque sólo el 22% aproximadamente de los usuarios consultan sobre el promedio, la mayoría de ellos (14% del total), tienen notas por sobre el promedio.

Tabla 8-3: Uso de colección electrónica y notas 2012-2014

2012-2014 Uso de colección electrónica y notas			
Promedios	**Zscore <0,2**	**Zscore >=0,2**	**Total x bca**
Sesiones >=181	35.17%	64.83%	3472
Sesiones <181	46.99%	53.01%	12976
Total x zscore	44.50%	55.50%	16448
X-squared = 154.67, df = 1, p-value < 2.2e-16			

El p-value obtenido es 2.2e-16, menor a 0.05, lo que indica que existe una diferencia significativa en el rendimiento académico de quienes usan la colección electrónica por sobre el promedio, lo que permite rechazar la hipótesis nula.

Tabla 8-4: Duración de sesiones y notas 2012-2014

2012-2014 Duración de sesiones y notas			
Promedios	**Zscore <0.2**	**Zscore>=0.2**	**Total**
Duración >437	39,49%	60,51%	4877
Duración <437	46,61%	53,39%	11571
Total x zscore	44,50%	55,50%	16448
X-squared = 67.731, df = 1, p-value < 2.2e-16			

En los clusters obtenidos se observa que en los años 2012 a 2014 se detectó un grupo menor de usuarios que tenía la mayor cantidad de transacciones en la biblioteca cada año, el patrón común de estos grupos es que eran alumnos con notas por sobre el promedio.

En los clusters obtenidos se observa en la relación notas-sesiones y notas-duración de cada año, un grupo pequeño de usuarios que presentan la mayor cantidad de sesiones y la duración de sesiones más alta, el patrón común de estos grupos es que son alumnos con notas por sobre el promedio.

Como se ha mencionado anteriormente, tanto para la colección presencial como para la electrónica, no se observan casos de alumnos con bajas notas que tengan un alto uso de

los servicios de la biblioteca. Se observan si alumnos con alto rendimiento que hacen poco uso de la biblioteca y sus colecciones.

Por lo tanto y como una tendencia, mayor para la colección presencial y leve para la colección electrónica, se corrobora la hipótesis de que los alumnos con mejor rendimiento son los que consultan más los servicios de la biblioteca.

Hipótesis 3:

H_0: No existe relación entre las consultas a la colección electrónica y el uso de la biblioteca presencial.

H_1: Existe relación entre las consultas a la colección electrónica y el uso de la biblioteca presencial

La correlación entre las variables de consultas a la colección electrónica y uso de la biblioteca es positiva, aunque muy leve, cercano a cero. Se observó un aumento de esta correlación entre 2012 al 2014, un 1% aproximadamente cada año, así como del tamaño del efecto de una variable sobre la otra.

Tabla 8-5: Correlación entre Consultas presenciales y a la colección electrónica.

Estadísticas	2012-2014
Coeficiente de correlación múltiple	0.057154238
Tamaño del efecto	0.003266607
R^2 ajustado	0.003228648
Error típico	828.5097486
Observaciones	26260

Existe una leve tendencia a la disminución de las transacciones presenciales de los usuarios en la biblioteca, mientras que las consultas a la colección electrónica van en aumento. Sin embargo una tendencia no va en detrimento de la otra, ya que se complementan durante los años de carrera de los usuarios, una baja mientras la otra sube en los últimos años, pero hay un periodo intermedio en que ambas subsisten.

Por lo tanto se rechaza la hipótesis nula que indicaba que no existe relación entre las consultas a la colección electrónica y las consultas de material impreso. Existe una relación, aunque muy baja.

8.4 Sugerencias

Buscando dar una aplicación práctica a los resultados del estudio realizado, se entregan las siguientes sugerencias y planteamientos de trabajos a futuro con el fin de apoyar la mejora y desarrollo de servicios en Bibliotecas UC.

Formación de usuarios

- Promover los diversos servicios de la biblioteca entre las carreras de las áreas que tienen consulta por debajo del promedio. (áreas de Ciencia y Tecnología, Cs. Médicas y Cs. Sociales).
- Promover el uso de la colección electrónica en talleres impartidos a partir de los primeros años de carrera.

Estudio de usuarios

- Realizar un estudio de usuarios en las áreas con menos uso de la colección presencial y electrónica, para profundizar en las razones del bajo uso de biblioteca que presentan.

Difusión

- Promover el uso de la biblioteca entre los alumnos, en página web y cartelería electrónica, difundiendo la asociación de uso de biblioteca con notas altas,
- Promover en la interfaz de los servicios electrónicos el uso de los servicios presenciales de la biblioteca, y en los sitios de los servicios presenciales promover el uso de la colección electrónica.
- Difundir entre los profesores la asociación de uso de biblioteca con notas altas, para promover la participación en la actualización de las bibliografías mínimas.

Desarrollo de colecciones

- Promover la inclusión de recursos electrónicos en las bibliografías mínimas.
- Impulsar la adquisición de nuevos materiales que sean de interés en las áreas con menos uso.

Gestión

- Incorporar nuevas herramientas para análisis y visualización de datos en los procesos de gestión estadística de la biblioteca (SPSS, R, Rapid Miner, etc.).
- Incorporar operadores y modelos de minería de datos en los análisis de datos de la biblioteca, con el objetivo de ir descubriendo patrones de uso de los diversos servicios.
- Explotar y establecer análisis regulares de los logs obtenidos desde el servidor Ezproxy, ya que es una valiosa fuente de datos.
- Establecer una fuente única, base de datos, que permita recolectar e integrar los datos de los distintos servicios para optimizar el análisis de ellos.
- Capacitar al personal en las nuevas herramientas de análisis y visualización de los datos.
- Considerar el desarrollo de una aplicación que permita realizar recomendaciones a partir de las consultas de nuestros usuarios, con el fin de promover las consultas multidisciplinarias.

Es importante destacar que no se puede limitar el valor de la biblioteca dentro de la Universidad a la nota obtenida por los estudiantes que acuden a usar sus servicios, es necesario continuar con estudios respecto al impacto de la biblioteca en distintos ámbitos del quehacer universitario y la percepción que tienen los usuarios de ella.

Al finalizar este trabajo, se tiene la certeza de haber enriquecido la experiencia profesional con esta metodología antes desconocida, que puede aportar valiosos análisis para la gestión de Bibliotecas UC. Se espera continuar la línea de estudio en esta área, considerando que aún existen diversos datos que pueden ser explotados de una forma óptima.

Bibliografía consultada

1. AFIONTZI, E., KAZADEIS, G., PAPACHRISTOPOULOS, L., SFAKAKIS, M., TSAKONAS, G. y PAPATHEODOROU, C. (2013) Semantically Enhanced Mining Methodology. In *Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries*. ACM, Indiana.
2. ALVARADO, L. (2015) *Gestión Integral de proyectos: taller proyecto de tesis*. 2 v.
3. BREEDING, M. (2013) Mining data for library decision support. *Computers in Libraries*, 33(5), 23-25.
   Retrieved from http://ezproxy.puc.cl/docview/1372419731?accountid=16788
4. BUCHEL, O. (2011) Designing Map-Based Visualizations for Collection Understanding. In Proceedings of the 11th Annual Joint Conference on Digital Libraries. Ottawa, ACM, pp.429
5. CHANG, C. y CHEN, R. (2006) Using data mining technology to solve classification problems. *The Electronic Library*,24(3), 307-321.
   doi:http://dx.doi.org/10.1108/02640470610671178
6. CHEN, Y., LIU, Y., y HO, W. (2013) A text mining approach to assist the general public in the retrieval of legal documents. *Journal of the American Society for Information Science and Technology*, 64(2), 280.
   Recuperado desde: http://ezproxy.puc.cl/docview/1277972107?accountid=16788
7. CHOI, Y. (2013) Analysis of image search queries on the web: Query modification patterns and semantic attributes. *Journal of the American Society for Information Science and Technology*, 64(7), 1423.
   Retrieved from http://ezproxy.puc.cl/docview/1365656608?accountid=16788
8. COX, B. L., & JANTTI, M. (2012) Capturing business intelligence required for targeted marketing, demonstrating value, and driving process improvement. *Library & Information Science Research*, *34*(4), 308–316.
   http://doi.org/10.1016/j.lisr.2012.06.002
9. DENG, H., HAN, J., LYU, M. y KING, I. (2012) Networks for Expertise Ranking. In Proceedings of the 12th Annual Joint Conference on Digital Libraries. Washington, ACM, pp.71
10. DOLAMIC, L., y SAVOY, J. (2010) Retrieval effectiveness of machine translated queries. *Journal of the American Society for Information Science and Technology*, 61(11), 2266.

Recuperado desde: http://ezproxy.puc.cl/docview/758864740?accountid=16788

11. FAYYAD, U., PIATETSKY-SHAPIRO, G. Y SMITH, P. (1996) From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, 17(3), 37-54.
12. GIBSON, I. E. (2001) *Data mining analysis of digital library database usage patterns as a tool facilitating efficient user navigation* (Order No. 3038905). Available from ProQuest Dissertations & Theses Global. (251131313).
    Retrieved from http://ezproxy.puc.cl/docview/251131313?accountid=16788
13. GOODALL, D., & PATTERN, D. (2011) Academic library non/low use and undergraduate student achievement. *Library Management*, *32*(3), 159–170. http://doi.org/10.1108/01435121111112871
14. GUENTHER, K. (2000) Applying data mining principles to library data collection. *Computers in Libraries*, 20(4), 60-63.
    Recuperado desde: http://ezproxy.puc.cl/docview/231142047?accountid=16788
15. HINZE, A., MCKAY, D. y VANDERSCHANTZ, N. (2012) Book Selection Behavior in the Physical Library: Implications for eBook Collections. In Proceedings of the 12th Annual Joint Conference on Digital Libraries. Washington, ACM, pp.305
16. KAWASE, R. y HERDER, E. (2011) Classification of User Interest Patterns Using a Virtual Folksonomy. In Proceedings of the 11th Annual Joint Conference on Digital Libraries. Ottawa, ACM, pp.105
17. JARA, M. et al. (2013) Proyecto Fondecyt 1131109: "Learning analytics: understanding relationships between eLearning and library use, student perceptions of these tools and learning outcomes". Documento de postulación del proyecto. Pontificia Universidad Católica de Chile.
18. KU, Y., CHIU, C., ZHANG, Y., CHEN, H., y SU, H. (2014) Text mining self-disclosing health information for public health service. *Journal of the Association for Information Science and Technology*, 65(5), 928.
    Recuperado desde: http://ezproxy.puc.cl/docview/1518990324?accountid=16788
19. LESKOVEC, J., RAJARAMAN, A. y ULLMAN, J. (2014) Mining of massive datasets. Stanford University.
20. LIN, Y., LIN, H., XU, K., y SUN, X. (2013) Learning to rank using smoothing methods for language modeling. *Journal of the American Society for Information Science and Technology*, 64(4), 818.
    Retrieved from http://ezproxy.puc.cl/docview/1319255789?accountid=16788

21. LIU, X., & TURTLE, H. (2013) Real-time user interest modeling for real-time ranking. *Journal of the American Society for Information Science and Technology*, 64(8), 1557. Recuperado desde: http://ezproxy.puc.cl/docview/1411811628?accountid=16788
22. NEWMAN, D., NOH, Y., TALLEY, E., KARIMI, S. y BALDWIN, T. (2010) Evaluating Topic Models for Digital Libraries. In Proceedings of the 11th Annual Joint Conference on Digital Libraries. Queensland, Australia, ACM, pp.215
23. NICHOLSON, S. (2003) The bibliomining process: Data warehousing and data mining for library decision making. *Information Technology and Libraries*, 22(4), 146-151. Recuperado desde: http://ezproxy.puc.cl/docview/215828522?accountid=16788
24. NICHOLSON, S. y STANTON, J. (2003) Gaining strategi advantage through bibliomining: data mining for management decisions in corporate, special, digital, and traditional libraries. In H. Nemati y C. Barko (eds.). *Organizational data mining: leveraging enterprise data resources for optimal performance*. Idea Group Pub., Hershey, Penn.
25. NIE, M., y ZHOU, X. (2012) Study on library administrate decision using data mining. *International Journal of Advancements in Computing Technology*, 4(8), 165-172. Retrieved from http://ezproxy.puc.cl/docview/1018556962?accountid=16788
26. PALMER, B. C. (2012) *Web usage mining: Application to an online educational digital library service* (Order No. 3507476). Available from ProQuest Dissertations & Theses Global. (1015628733). Recuperado desde: http://ezproxy.puc.cl/docview/1015628733?accountid=16788
27. PALMER, C., WEBER, N. y CRAGIN, M. (2011) Analytic Potential of Data: Assessing Reuse Value. In Proceedings of the 11th Annual Joint Conference on Digital Libraries. Ottawa, ACM, pp.425
28. RODRÍGUEZ-BÁRCENAS, G., y LÓPEZ-HUERTAS, M., J. (2013) Saaty's analytic hierarchies method for knowledge organization in decision making. *Journal of the American Society for Information Science and Technology*, 64(7), 1454. Recuperado desde: http://ezproxy.puc.cl/docview/1365656613?accountid=16788
29. SORIA, K., FRANSEN, J., y NACKERUD, S. (2013) Library Use and Undergraduate Student Outcomes: New Evidence for Students' Retention and Academic Success. Recuperado desde: http://conservancy.umn.edu/handle/11299/143312

30. STONE, G., y RAMSDEN, B. (2013) Library Impact Data Project: Looking for the Link between Library Usage and Student Attainment. College & Research Libraries, 74(6), 546–559. http://doi.org/10.5860/crl12-406
31. SUN, X., y LIN, H. (2013) Topical community detection from mining user tagging behavior and interest. *Journal of the American Society for Information Science and Technology*, 64(2), 321.
Retrieved from http://ezproxy.puc.cl/docview/1277971966?accountid=16788
32. WILSON, M. J., y WILSON, M. L. (2013) A comparison of techniques for measuring sensemaking and learning within participant-generated summaries. *Journal of the American Society for Information Science and Technology*, 64(2), 291. Recuperado desde: http://ezproxy.puc.cl/docview/1277971939?accountid=16788
33. XU, B., y RECKER, M. (2012) Teaching analytics: A clustering and triangulation study of digital library user data. *Journal of Educational Technology & Society*, 15(3), 103-n/a.
Recuperado desde: http://ezproxy.puc.cl/docview/1287025343?accountid=16788
34. Wikipedia: la encyclopedia libre. Consultada durante febrero 2016 en: https://es.wikipedia.org/w/index.php?
35. WONG, S. H. R., & WEBB, T. D. (2010) Uncovering Meaningful Correlation between Student Academic Performance and Library Material Usage. College & Research Libraries, 72(4), 361–370. http://doi.org/10.5860/crl-129
36. WYNHOLDS, L., WALLIS, J., BORGMAN, C., SANDS, A. y TRAWEEK, S. (2012) Data, Data Use, and Scientific Inquiry: Two Case Studies of Data Practices. In Proceedings of the 12th Annual Joint Conference on Digital Libraries. Washington, ACM, pp. 19
37. ZHENG, W., WANG, X., FANG, H. y CHENG, H. (2011) An Exploration of Pattern-based Subtopic Modeling for Search Result Diversification. In Proceedings of the 11th Annual Joint Conference on Digital Libraries. Ottawa, ACM, pp.385
38. ZHU, W. Z., y ALLEN, R. B. (2013) Document clustering using the LSI subspace signature model. *Journal of the American Society for Information Science and Technology*, 64(4), 844.
Recuperado desde: http://ezproxy.puc.cl/docview/1319255784?accountid=1678.

Anexos

Anexo 1: Consultas SQL utilizadas en tablas de Aleph

Query 1: Transacciones por año, con áreas temáticas y tipo de usuario.

```
SELECT "PUC50"."V_Z35"."Z35_EVENT_DATE",
        "PUC50"."V_Z13_CAR"."Z13_TITLE",
        "PUC50"."V_Z30"."Z30_INVENTORY_NUMBER",
        "PUC50"."V_Z30"."Z30_CALL_NO",
        "PUC50"."V_Z30"."Z30_MATERIAL",
        "PUC50"."V_Z35"."Z35_ID",
        "PUC50"."V_Z35"."Z35_BOR_STATUS",
        "PUC50"."V_Z35"."Z35_BOR_TYPE"
   FROM "PUC50"."V_Z13_CAR",
        "PUC50"."V_Z30",
        "PUC50"."V_Z304",
        "PUC50"."V_Z35"
  WHERE ( "PUC50"."V_Z13_CAR"."Z13_DOC_NUMBER" =
"PUC50"."V_Z30"."Z30_DOC_NUMBER" ) and
        ( "PUC50"."V_Z30"."Z30_DOC_NUMBER" =
"PUC50"."V_Z35"."Z35_DOC_NUMBER" ) and
        ( "PUC50"."V_Z35"."Z35_ITEM_SEQUENCE" =
"PUC50"."V_Z30"."Z30_ITEM_SEQUENCE" ) and
        ( "PUC50"."V_Z35"."Z35_ID" = "PUC50"."V_Z304"."Z304_ID" ) and
        ( ( "PUC50"."V_Z35"."Z35_EVENT_TYPE" in (‘50’, '62', '63') ) AND
        ( "PUC50"."V_Z35"."Z35_EVENT_DATE" BETWEEN  '20140101' AND '20141231' )
AND  ( "PUC50"."V_Z30"."Z30_MATERIAL"  not in  ('COMPU', 'SEMI', 'AUDIF', 'CUBI',
'LOCKE', 'TVDVD', 'TVVHS', 'TES', 'ISSUE', 'ISBD’)  )   )
ORDER BY "PUC50"."V_Z35"."Z35_EVENT_DATE" ASC   ;
```

Anexo 2: Tipos de usuarios definidos en la Biblioteca

CODIGO	DESCRIPCION
1	Alumno Egresado
2	Alumno Interno Música
3	Alumno Regular
4	Alumno Regular 2
5	Alumno Provisional
6	Alumno Cursos de Int.
20	Alumno en vías de Titulación
21	Alumno Vigente Egresado
23	Alumno PIANE
30	Alumno Especial Biblioteca Teología
7	Docente
8	Administrativo
9	Convenio
10	Profesor Emerito
11	Investigador Bibliog.
12	Institucion
13	PI/2
14	Otros
15	Intercampus
16	Profesor Retirado
17	Seccion Reserva
18	Usuario sin Privilegios
19	Préstamo Interbibliotecario Internacional
22	Institución - 14 días
24	Postdoctorado
25	Proyecto Biblioteca Escolar
26	Proyecto Cededoc
27	Administrativo Retirado
28	Alumno B. Valle del Elqui
29	Docente B. Valle del Elqui

Anexo 3: Áreas temáticas y Áreas OCDE

CODIGO	DESCRIPCION	OCDE	General
0	GENERALIDADES	Cs. Sociales	000
1	GENERALIDADES	Cs. Sociales	001
2	GENERALIDADES	Cs. Sociales	002
3	COMPUTACION	Cs. Naturales	003
4	COMPUTACION	Cs. Naturales	004
5	COMPUTACION	Cs. Naturales	005
6	COMPUTACION	Cs. Naturales	006
10	BIBLIOGRAFIAS	Cs. Sociales	010
20	BIBLIOTECOLOGIA	Cs. Sociales	020
30	OBRAS ENCICLOPEDICAS GENERALES	Cs. Sociales	030
50	PUBLICACIONES SERIADAS GENERALES	Cs. Sociales	050
60	ORGANIZACIONES GENERALES Y MUSEOLOGIA	Cs. Sociales	060
70	PERIODISMO	Cs. Sociales	070
80	COLECCIONES	Cs. Sociales	080
90	MANUSCRITOS Y LIBROS RAROS	Cs. Sociales	090
100	FILOSOFIA	Humanidades	100
110	METAFISICA	Humanidades	110
11185	BELLEZA	Humanidades	110
120	EPISTEMOLOGIA	Humanidades	120
130	PARAPSICOLOGIA Y OCULTISMO	Humanidades	130
140	ESCUELAS DE FILOSOFIA (PLATONISMO)	Humanidades	140
150	PSICOLOGIA	Cs. Sociales	150
155	PSICOLOGIA, ENFERMERIA	Cs. Sociales	150
160	LOGICA	Humanidades	160
170	ETICA	Humanidades	170
1743	ETICA DE PROFESIONES LEGALES	Humanidades	170
180	FILOSOFIA ANTIGUA, MEDIEVAL Y ORIENTAL	Humanidades	180
190	FILOSOFIA MODERNA OCCIDENTAL	Humanidades	190
200	RELIGION	Humanidades	200
2629	DERECHO ECLESIASTICO	Humanidades	260
300	SOCIOLOGIA	Cs. Sociales	300
300	TRABAJO SOCIAL	Cs. Sociales	300
301	SOCIOLOGIA-ANTROPOLOGIA	Cs. Sociales	300
3022	COMUNICACIONES	Cs. Sociales	300
3022	SOCIOLOGIA	Cs. Sociales	300
3023	DINAMICAS DE GRUPO	Cs. Sociales	300
3023	INTERACCION SOCIAL EN GRUPOS	Cs. Sociales	300
307	COMUNIDADES	Cs. Sociales	300
310	ESTADISTICA	Cs. Sociales	310
320	CIENCIA POLITICA	Cs. Sociales	320
330	ECONOMIA	Cs. Sociales	330
3337	RECURSOS NATURALES	Cs. Sociales	330
3381	INDUSTRIA AGRICOLA	Cs. Sociales	330
340	DERECHO	Cs. Sociales	340
350	ADMINISTRACION PUBLICA	Cs. Sociales	350
351	ADMINISTRACION PUBLICA	Cs. Sociales	350
352	ADMINISTRACION PUBLICA	Cs. Sociales	350
353	ADMINISTRACION PUBLICA	Cs. Sociales	350
354	ADMINISTRACION PUBLICA	Cs. Sociales	350
355	CIENCIA MILITAR	Cs. Sociales	350
356	CIENCIA MILITAR	Cs. Sociales	350
357	CIENCIA MILITAR	Cs. Sociales	350
358	CIENCIA MILITAR	Cs. Sociales	350
359	CIENCIA MILITAR	Cs. Sociales	350
360	PROBLEMAS SOCIALES Y TRABAJO SOCIAL	Cs. Sociales	360

3622	DESORDENES Y ENFERMEDADES EMOCIONALES ...	Cs. Sociales	360
3627	SERVICIOS Y PROBLEMAS DE JOVENES	Cs. Sociales	360
36331	CENSURA	Cs. Sociales	360
3638	NUTRICION Y DIETETICA	Cs. Sociales	360
364	CRIMINOLOGIA	Cs. Sociales	360
370	EDUCACION	Cs. Sociales	370
380	COMERCIO, COMUNICACIÓN Y TRANSPORTE	Cs. Sociales	380
382	COMERCIO INTERNACIONAL	Cs. Sociales	380
390	COSTUMBRES Y FOLKLORE	Cs. Sociales	390
400	LENGUAJE	Humanidades	400
500	CIENCIAS NATURALES Y MATEMATICAS	Cs. Naturales	500
500	CIENCIA	Cs. Naturales	500
508	HISTORIA NATURAL	Cs. Naturales	500
509	GENERALES	Cs. Naturales	500
510	MATEMATICAS	Cs. Naturales	510
520	ASTRONOMIA	Cs. Naturales	520
530	FISICA	Cs. Naturales	530
530	ASTRONOMIA	Cs. Naturales	530
540	QUIMICA	Cs. Naturales	540
550	GEOLOGIA	Cs. Naturales	550
550	CIENCIAS DE LA TIERRA	Cs. Naturales	550
560	PALEONTOLOGIA Y PALEOZOOLOGIA	Cs. Naturales	560
570	BIOLOGIA	Cs. Naturales	570
577	ECOLOGIA	Cs. Naturales	570
580	BOTANICA	Cs. Naturales	580
590	ZOOLOGIA	Cs. Naturales	590
600	TECNOLOGIA	Ingeniería y Tecnología	600
610	MEDICINA Y CIENCIAS MEDICAS	Cs. Médicas y de la Salud	610
6141	MEDICINA Y CIENCIAS MEDICAS	Cs. Médicas y de la Salud	610
61689	MEDICINA Y CIENCIAS MEDICAS	Cs. Médicas y de la Salud	610
6178	MEDICINA Y CIENCIAS MEDICAS	Cs. Médicas y de la Salud	610
6137	MEDICINA Y CIENCIAS MEDICAS	Cs. Médicas y de la Salud	610
6132	MEDICINA Y CIENCIAS MEDICAS	Cs. Médicas y de la Salud	610
6176	MEDICINA Y CIENCIAS MEDICAS	Cs. Médicas y de la Salud	610
6179676	MEDICINA Y CIENCIAS MEDICAS	Cs. Médicas y de la Salud	610
620	INGENIERIA	Ingeniería y Tecnología	620
628	INGENIERIA AMBIENTAL	Ingeniería y Tecnología	620
630	AGRICULTURA	Ingeniería y Tecnología	630
640	ECONOMIA DOMESTICA	Ingeniería y Tecnología	640
64122	VINO	Ingeniería y Tecnología	640
650	RELACIONES PUBLICAS Y ADMINISTRACION	Cs. Sociales	650
657	CONTABILIDAD	Cs. Sociales	650
658	ADMINISTRACION GENERAL	Cs. Sociales	650
6583	ADMINISTRACION DE RECURSO HUMANO	Cs. Sociales	650
659	PUBLICIDAD	Cs. Sociales	650
660	INGENIERIA QUIMICA	Ingeniería y Tecnología	660
669	METALURGIA	Ingeniería y Tecnología	660
670	MANUFACTURAS	Ingeniería y Tecnología	670
6794	MATERIALES DENTALES	Ingeniería y Tecnología	670
680	MANUFACTURAS	Ingeniería y Tecnología	680
68622	TIPOGRAFIA	Ingeniería y Tecnología	680
690	CONSTRUCCION	Ingeniería y Tecnología	690
700	ARTES	Humanidades	700
7001	FILOSOFÍA DE LAS BELLAS ARTES	Humanidades	700
710	PAISAJISMO Y PLANIFICACION URBANA	Humanidades	710
720	ARQUITECTURA	Humanidades	720
730	ESCULTURA	Humanidades	730
740	DIBUJO Y ARTES DECORATIVAS	Humanidades	740
750	PINTURA	Humanidades	750

760	ARTES GRAFICAS	Humanidades	760
770	FOTOGRAFIA Y ARTE POR COMPUTADOR	Humanidades	770
780	MUSICA	Humanidades	780
790	DEPORTES, JUEGOS Y ENTRETENIMIENTOS	Humanidades	790
7914	PELICULAS, RADIO, TELEVISION	Humanidades	790
792	TEATRO	Humanidades	790
794	JUEGOS DE INTERIOR	Humanidades	790
795	JUEGOS DE PROBABILIDAD	Humanidades	790
796	ATLETISMO Y JUEGOS DE EXTERIOR	Humanidades	790
797	DEPORTES ACUATICOS Y AEREOS	Humanidades	790
798	DEPORTES ECUESTRES Y CARRERS DE ANIMALES	Humanidades	790
799	PESCA, CAZA, TIRO	Humanidades	790
800	LITERATURA, RETORICA Y CRITICA	Humanidades	800
80193	ESTÉTICA	Humanidades	800
900	HISTORIA	Humanidades	900
910	GEOGRAFIA	Humanidades	910
920	BIOGRAFIA Y GENEALOGIA	Humanidades	920
930	HISTORIA	Humanidades	930
940	HISTORIA	Humanidades	940
950	HISTORIA	Humanidades	950
960	HISTORIA	Humanidades	960
970	HISTORIA	Humanidades	970
980	HISTORIA	Humanidades	980
990	HISTORIA	Humanidades	990

Anexo 4: Reglas de asociación 2012-2014

2012 – Reglas de Asociación (100 primeras)

No.	Premisa	Conclusión	Soporte	Confianza
161	content.ebscohost.com	web.ebscohost.com	0,63	1,00
162	media.proquest.com	search.proquest.com	0,46	1,00
164	www.sciencedirect.com, content.ebscohost.com	web.ebscohost.com	0,44	1,00
165	content.ebscohost.com, search.proquest.com	web.ebscohost.com	0,44	1,00
170	web.ebscohost.com, media.proquest.com	search.proquest.com	0,40	1,00
166	content.ebscohost.com, media.proquest.com	web.ebscohost.com	0,36	1,00
174	content.ebscohost.com, media.proquest.com	search.proquest.com	0,36	1,00
183	content.ebscohost.com, media.proquest.com	web.ebscohost.com, search.proquest.com	0,36	1,00
184	web.ebscohost.com, content.ebscohost.com, media.proquest.com	search.proquest.com	0,36	1,00
185	content.ebscohost.com, search.proquest.com, media.proquest.com	web.ebscohost.com	0,36	1,00
172	www.sciencedirect.com, media.proquest.com	search.proquest.com	0,34	1,00
177	www.sciencedirect.com, content.ebscohost.com, search.proquest.com	web.ebscohost.com	0,33	1,00
163	static.pubmed.gov	www.ncbi.nlm.nih.gov	0,32	1,00
181	web.ebscohost.com, www.sciencedirect.com, media.proquest.com	search.proquest.com	0,31	1,00
168	content.ebscohost.com, openurl.ebscohost.com	web.ebscohost.com	0,29	1,00
178	www.sciencedirect.com, content.ebscohost.com, media.proquest.com	web.ebscohost.com	0,28	1,00
191	www.sciencedirect.com, content.ebscohost.com, media.proquest.com	search.proquest.com	0,28	1,00
195	www.sciencedirect.com, content.ebscohost.com, media.proquest.com	web.ebscohost.com, search.proquest.com	0,28	1,00
196	web.ebscohost.com, www.sciencedirect.com, content.ebscohost.com, media.proquest.com	search.proquest.com	0,28	1,00
197	www.sciencedirect.com, content.ebscohost.com, search.proquest.com, media.proquest.com	web.ebscohost.com	0,28	1,00
167	content.ebscohost.com, www.springerlink.com	web.ebscohost.com	0,27	1,00
173	www.sciencedirect.com, static.pubmed.gov	www.ncbi.nlm.nih.gov	0,26	1,00
179	www.sciencedirect.com, content.ebscohost.com, www.springerlink.com	web.ebscohost.com	0,25	1,00
180	www.sciencedirect.com, content.ebscohost.com, openurl.ebscohost.com	web.ebscohost.com	0,25	1,00
187	content.ebscohost.com, search.proquest.com, openurl.ebscohost.com	web.ebscohost.com	0,24	1,00
176	media.proquest.com, openurl.ebscohost.com	search.proquest.com	0,24	1,00
171	web.ebscohost.com, static.pubmed.gov	www.ncbi.nlm.nih.gov	0,22	1,00
186	content.ebscohost.com, search.proquest.com, www.springerlink.com	web.ebscohost.com	0,22	1,00
190	web.ebscohost.com, media.proquest.com, openurl.ebscohost.com	search.proquest.com	0,22	1,00
175	media.proquest.com, www.springerlink.com	search.proquest.com	0,22	1,00
199	www.sciencedirect.com, content.ebscohost.com, search.proquest.com, openurl.ebscohost.com	web.ebscohost.com	0,21	1,00
198	www.sciencedirect.com, content.ebscohost.com, search.proquest.com, www.springerlink.com	web.ebscohost.com	0,21	1,00
189	web.ebscohost.com, media.proquest.com, www.springerlink.com	search.proquest.com	0,21	1,00

188	content.ebscohost.com, media.proquest.com, openurl.ebscohost.com	web.ebscohost.com	0,21	1,00
194	content.ebscohost.com, media.proquest.com, openurl.ebscohost.com	search.proquest.com	0,21	1,00
200	content.ebscohost.com, media.proquest.com, openurl.ebscohost.com	web.ebscohost.com, search.proquest.com	0,21	1,00
201	web.ebscohost.com, content.ebscohost.com, media.proquest.com, openurl.ebscohost.com	search.proquest.com	0,21	1,00
202	content.ebscohost.com, search.proquest.com, media.proquest.com, openurl.ebscohost.com	web.ebscohost.com	0,21	1,00
192	www.sciencedirect.com, media.proquest.com, www.springerlink.com	search.proquest.com	0,21	1,00
169	content.ebscohost.com, onlinelibrary.wiley.com	web.ebscohost.com	0,21	1,00
193	www.sciencedirect.com, media.proquest.com, openurl.ebscohost.com	search.proquest.com	0,21	1,00
182	web.ebscohost.com, www.sciencedirect.com, static.pubmed.gov	www.ncbi.nlm.nih.gov	0,20	1,00
160	resources.metapress.com	www.springerlink.com	0,24	1,00
159	www.sciencedirect.com, resources.metapress.com	www.springerlink.com	0,22	1,00
158	www.ncbi.nlm.nih.gov	static.pubmed.gov	0,32	0,98
157	web.ebscohost.com, www.ncbi.nlm.nih.gov	static.pubmed.gov	0,22	0,98
156	www.sciencedirect.com, www.ncbi.nlm.nih.gov	static.pubmed.gov	0,26	0,98
155	web.ebscohost.com, www.sciencedirect.com, www.ncbi.nlm.nih.gov	static.pubmed.gov	0,20	0,98
154	www.sciencedirect.com, search.proquest.com, openurl.ebscohost.com	web.ebscohost.com	0,23	0,94
151	media.proquest.com, openurl.ebscohost.com	web.ebscohost.com	0,22	0,94
152	media.proquest.com, openurl.ebscohost.com	web.ebscohost.com, search.proquest.com	0,22	0,94
153	search.proquest.com, media.proquest.com, openurl.ebscohost.com	web.ebscohost.com	0,22	0,94
150	search.proquest.com, openurl.ebscohost.com	web.ebscohost.com	0,26	0,94
147	media.proquest.com, www.springerlink.com	web.ebscohost.com	0,21	0,94
148	media.proquest.com, www.springerlink.com	web.ebscohost.com, search.proquest.com	0,21	0,94
149	search.proquest.com, media.proquest.com, www.springerlink.com	web.ebscohost.com	0,21	0,94
144	content.ebscohost.com, search.proquest.com, www.springerlink.com	www.sciencedirect.com	0,21	0,93
145	content.ebscohost.com, search.proquest.com, www.springerlink.com	web.ebscohost.com, www.sciencedirect.com	0,21	0,93
146	web.ebscohost.com, content.ebscohost.com, search.proquest.com, www.springerlink.com	www.sciencedirect.com	0,21	0,93
141	media.proquest.com, www.springerlink.com	www.sciencedirect.com	0,21	0,93
142	media.proquest.com, www.springerlink.com	www.sciencedirect.com, search.proquest.com	0,21	0,93
143	search.proquest.com, media.proquest.com, www.springerlink.com	www.sciencedirect.com	0,21	0,93
140	web.ebscohost.com, search.proquest.com, www.springerlink.com	www.sciencedirect.com	0,23	0,93
137	web.ebscohost.com, media.proquest.com, openurl.ebscohost.com	content.ebscohost.com	0,21	0,93
138	web.ebscohost.com, media.proquest.com, openurl.ebscohost.com	content.ebscohost.com, search.proquest.com	0,21	0,93
139	web.ebscohost.com, search.proquest.com, media.proquest.com, openurl.ebscohost.com	content.ebscohost.com	0,21	0,93
136	ovidsp.tx.ovid.com	www.sciencedirect.com	0,20	0,93
135	search.proquest.com, www.springerlink.com	www.sciencedirect.com	0,25	0,93

134	web.ebscohost.com, www.sciencedirect.com, search.proquest.com, openurl.ebscohost.com	content.ebscohost.com	0,21	0,93
131	content.ebscohost.com, www.springerlink.com	www.sciencedirect.com	0,25	0,92
132	content.ebscohost.com, www.springerlink.com	web.ebscohost.com, www.sciencedirect.com	0,25	0,92
133	web.ebscohost.com, content.ebscohost.com, www.springerlink.com	www.sciencedirect.com	0,25	0,92
130	web.ebscohost.com, onlinelibrary.wiley.com	www.sciencedirect.com	0,22	0,92
129	www.sciencedirect.com, search.proquest.com, www.springerlink.com	web.ebscohost.com	0,23	0,92
128	web.ebscohost.com, www.springerlink.com	www.sciencedirect.com	0,29	0,92
127	web.ebscohost.com, search.proquest.com, openurl.ebscohost.com	content.ebscohost.com	0,24	0,92
126	search.proquest.com, www.springerlink.com	web.ebscohost.com	0,25	0,92
125	resources.metapress.com	www.sciencedirect.com	0,22	0,92
124	www.springerlink.com, resources.metapress.com	www.sciencedirect.com	0,22	0,92
123	web.ebscohost.com, www.sciencedirect.com, openurl.ebscohost.com	content.ebscohost.com	0,25	0,91
122	onlinelibrary.wiley.com	www.sciencedirect.com	0,27	0,91
121	resources.metapress.com	www.sciencedirect.com, www.springerlink.com	0,22	0,91
120	www.sciencedirect.com, openurl.ebscohost.com	web.ebscohost.com	0,27	0,91
119	web.ebscohost.com, openurl.ebscohost.com	content.ebscohost.com	0,29	0,91
118	www.springerlink.com	www.sciencedirect.com	0,35	0,91
117	openurl.ebscohost.com	web.ebscohost.com	0,32	0,91
114	www.sciencedirect.com, media.proquest.com	web.ebscohost.com	0,31	0,91
115	www.sciencedirect.com, media.proquest.com	web.ebscohost.com, search.proquest.com	0,31	0,91
116	www.sciencedirect.com, search.proquest.com, media.proquest.com	web.ebscohost.com	0,31	0,91
111	web.ebscohost.com, www.sciencedirect.com, media.proquest.com	content.ebscohost.com	0,28	0,90
112	web.ebscohost.com, www.sciencedirect.com, media.proquest.com	content.ebscohost.com, search.proquest.com	0,28	0,90
113	web.ebscohost.com, www.sciencedirect.com, search.proquest.com, media.proquest.com	content.ebscohost.com	0,28	0,90
110	web.ebscohost.com, www.sciencedirect.com, search.proquest.com, www.springerlink.com	content.ebscohost.com	0,21	0,90
109	web.ebscohost.com, search.proquest.com, www.springerlink.com	content.ebscohost.com	0,22	0,89
106	web.ebscohost.com, media.proquest.com	content.ebscohost.com	0,36	0,89
107	web.ebscohost.com, media.proquest.com	content.ebscohost.com, search.proquest.com	0,36	0,89
108	web.ebscohost.com, search.proquest.com, media.proquest.com	content.ebscohost.com	0,36	0,89
105	www.sciencedirect.com, search.proquest.com	web.ebscohost.com	0,38	0,89
104	web.ebscohost.com, onlinelibrary.wiley.com	content.ebscohost.com	0,21	0,88

2013 – Reglas de Asociación (100 primeras)

No.	Premisa	Conclusión	Soporte	Confianza
156	media.proquest.com	search.proquest.com	0,43	1,00
161	search.proquest.com, content.ebscohost.com	web.ebscohost.com	0,37	1,00
162	web.ebscohost.com, media.proquest.com	search.proquest.com	0,37	1,00
159	www.sciencedirect.com, content.ebscohost.com	web.ebscohost.com	0,36	1,00
169	www.sciencedirect.com, media.proquest.com	search.proquest.com	0,32	1,00
163	content.ebscohost.com, media.proquest.com	web.ebscohost.com	0,30	1,00
173	content.ebscohost.com, media.proquest.com	search.proquest.com	0,30	1,00
185	content.ebscohost.com, media.proquest.com	web.ebscohost.com, search.proquest.com	0,30	1,00
186	web.ebscohost.com, content.ebscohost.com, media.proquest.com	search.proquest.com	0,30	1,00
187	search.proquest.com, content.ebscohost.com, media.proquest.com	web.ebscohost.com	0,30	1,00
178	www.sciencedirect.com, search.proquest.com, content.ebscohost.com	web.ebscohost.com	0,29	1,00
179	web.ebscohost.com, www.sciencedirect.com, media.proquest.com	search.proquest.com	0,28	1,00
157	rd.springer.com	link.springer.com	0,28	1,00
158	static.pubmed.gov	www.ncbi.nlm.nih.gov	0,27	1,00
170	www.sciencedirect.com, rd.springer.com	link.springer.com	0,26	1,00
155	ac.els-cdn.com	www.sciencedirect.com	0,25	1,00
180	www.sciencedirect.com, content.ebscohost.com, media.proquest.com	web.ebscohost.com	0,24	1,00
193	www.sciencedirect.com, content.ebscohost.com, media.proquest.com	search.proquest.com	0,24	1,00
197	www.sciencedirect.com, content.ebscohost.com, media.proquest.com	web.ebscohost.com, search.proquest.com	0,24	1,00
198	web.ebscohost.com, www.sciencedirect.com, content.ebscohost.com, media.proquest.com	search.proquest.com	0,24	1,00
199	www.sciencedirect.com, search.proquest.com, content.ebscohost.com, media.proquest.com	web.ebscohost.com	0,24	1,00
164	content.ebscohost.com, link.springer.com	web.ebscohost.com	0,23	1,00
171	www.sciencedirect.com, static.pubmed.gov	www.ncbi.nlm.nih.gov	0,22	1,00
172	ad.doubleclick.net, ac.els-cdn.com	www.sciencedirect.com	0,22	1,00
167	web.ebscohost.com, rd.springer.com	link.springer.com	0,22	1,00
181	www.sciencedirect.com, content.ebscohost.com, link.springer.com	web.ebscohost.com	0,22	1,00
166	content.ebscohost.com, openurl.ebscohost.com	web.ebscohost.com	0,22	1,00
174	media.proquest.com, link.springer.com	search.proquest.com	0,21	1,00
165	content.ebscohost.com, onlinelibrary.wiley.com	web.ebscohost.com	0,21	1,00
184	web.ebscohost.com, www.sciencedirect.com, rd.springer.com	link.springer.com	0,20	1,00
194	www.sciencedirect.com, media.proquest.com, link.springer.com	search.proquest.com	0,20	1,00
188	search.proquest.com, content.ebscohost.com, link.springer.com	web.ebscohost.com	0,20	1,00
177	search.proquest.com, rd.springer.com	link.springer.com	0,20	1,00
191	web.ebscohost.com, media.proquest.com, link.springer.com	search.proquest.com	0,20	1,00
182	www.sciencedirect.com, content.ebscohost.com, onlinelibrary.wiley.com	web.ebscohost.com	0,20	1,00
175	media.proquest.com, onlinelibrary.wiley.com	search.proquest.com	0,19	1,00
190	search.proquest.com, content.ebscohost.com, openurl.ebscohost.com	web.ebscohost.com	0,19	1,00

200	www.sciencedirect.com, search.proquest.com, content.ebscohost.com, link.springer.com	web.ebscohost.com	0,19	1,00
201	web.ebscohost.com, www.sciencedirect.com, media.proquest.com, link.springer.com	search.proquest.com	0,19	1,00
176	media.proquest.com, openurl.ebscohost.com	search.proquest.com	0,18	1,00
196	www.sciencedirect.com, search.proquest.com, rd.springer.com	link.springer.com	0,18	1,00
183	www.sciencedirect.com, content.ebscohost.com, openurl.ebscohost.com	web.ebscohost.com	0,18	1,00
160	web.ebscohost.com, ac.els-cdn.com	www.sciencedirect.com	0,18	1,00
168	web.ebscohost.com, static.pubmed.gov	www.ncbi.nlm.nih.gov	0,18	1,00
195	www.sciencedirect.com, media.proquest.com, onlinelibrary.wiley.com	search.proquest.com	0,18	1,00
192	web.ebscohost.com, search.proquest.com, rd.springer.com	link.springer.com	0,18	1,00
189	search.proquest.com, content.ebscohost.com, onlinelibrary.wiley.com	web.ebscohost.com	0,18	1,00
202	web.ebscohost.com, www.ncbi.nlm.nih.gov	static.pubmed.gov	0,18	0,99
154	www.ncbi.nlm.nih.gov	static.pubmed.gov	0,27	0,99
153	www.sciencedirect.com, www.ncbi.nlm.nih.gov	static.pubmed.gov	0,22	0,99
152	search.proquest.com, ad.doubleclick.net	www.sciencedirect.com	0,18	0,99
151	web.ebscohost.com, ad.doubleclick.net	www.sciencedirect.com	0,19	0,99
150	ad.doubleclick.net	www.sciencedirect.com	0,26	0,97
149	link.springer.com, onlinelibrary.wiley.com	www.sciencedirect.com	0,21	0,96
148	web.ebscohost.com, media.proquest.com, link.springer.com	www.sciencedirect.com	0,19	0,94
145	web.ebscohost.com, media.proquest.com, link.springer.com	www.sciencedirect.com, search.proquest.com	0,19	0,94
146	web.ebscohost.com, search.proquest.com, media.proquest.com, link.springer.com	www.sciencedirect.com	0,19	0,94
147	search.proquest.com, content.ebscohost.com, link.springer.com	www.sciencedirect.com	0,19	0,94
142	search.proquest.com, content.ebscohost.com, link.springer.com	web.ebscohost.com, www.sciencedirect.com	0,19	0,94
143	web.ebscohost.com, search.proquest.com, content.ebscohost.com, link.springer.com	www.sciencedirect.com	0,19	0,94
144	www.sciencedirect.com, search.proquest.com, openurl.ebscohost.com	web.ebscohost.com	0,19	0,94
141	media.proquest.com, link.springer.com	www.sciencedirect.com	0,20	0,94
138	media.proquest.com, link.springer.com	www.sciencedirect.com, search.proquest.com	0,20	0,94
139	search.proquest.com, media.proquest.com, link.springer.com	www.sciencedirect.com	0,20	0,94
140	search.proquest.com, openurl.ebscohost.com	web.ebscohost.com	0,22	0,94
137	web.ebscohost.com, search.proquest.com, link.springer.com	www.sciencedirect.com	0,23	0,93
136	web.ebscohost.com, search.proquest.com, onlinelibrary.wiley.com	www.sciencedirect.com	0,20	0,93
135	search.proquest.com, rd.springer.com	www.sciencedirect.com	0,18	0,93
132	search.proquest.com, rd.springer.com	www.sciencedirect.com, link.springer.com	0,18	0,93
133	search.proquest.com, link.springer.com, rd.springer.com	www.sciencedirect.com	0,18	0,93
134	content.ebscohost.com, link.springer.com	www.sciencedirect.com	0,22	0,93
129	content.ebscohost.com, link.springer.com	web.ebscohost.com, www.sciencedirect.com	0,22	0,93
130	web.ebscohost.com, content.ebscohost.com, link.springer.com	www.sciencedirect.com	0,22	0,93

131	search.proquest.com, link.springer.com	www.sciencedirect.com	0,25	0,93
128	media.proquest.com, onlinelibrary.wiley.com	www.sciencedirect.com	0,18	0,93
125	media.proquest.com, onlinelibrary.wiley.com	www.sciencedirect.com, search.proquest.com	0,18	0,93
126	search.proquest.com, media.proquest.com, onlinelibrary.wiley.com	www.sciencedirect.com	0,18	0,93
127	web.ebscohost.com, rd.springer.com	www.sciencedirect.com	0,20	0,93
122	web.ebscohost.com, rd.springer.com	www.sciencedirect.com, link.springer.com	0,20	0,93
123	web.ebscohost.com, link.springer.com, rd.springer.com	www.sciencedirect.com	0,20	0,93
124	ovidsp.tx.ovid.com	www.sciencedirect.com	0,20	0,92
121	www.sciencedirect.com, media.proquest.com, link.springer.com	web.ebscohost.com	0,19	0,92
118	www.sciencedirect.com, media.proquest.com, link.springer.com	web.ebscohost.com, search.proquest.com	0,19	0,92
119	www.sciencedirect.com, search.proquest.com, media.proquest.com, link.springer.com	web.ebscohost.com	0,19	0,92
120	content.ebscohost.com, onlinelibrary.wiley.com	www.sciencedirect.com	0,20	0,92
115	content.ebscohost.com, onlinelibrary.wiley.com	web.ebscohost.com, www.sciencedirect.com	0,20	0,92
116	web.ebscohost.com, content.ebscohost.com, onlinelibrary.wiley.com	www.sciencedirect.com	0,20	0,92
117	search.proquest.com, onlinelibrary.wiley.com	www.sciencedirect.com	0,22	0,92
114	web.ebscohost.com, link.springer.com	www.sciencedirect.com	0,27	0,92
113	web.ebscohost.com, onlinelibrary.wiley.com	www.sciencedirect.com	0,24	0,92
112	media.proquest.com, link.springer.com	web.ebscohost.com	0,20	0,92
109	media.proquest.com, link.springer.com	web.ebscohost.com, search.proquest.com	0,20	0,92
110	search.proquest.com, media.proquest.com, link.springer.com	web.ebscohost.com	0,20	0,92
111	rd.springer.com	www.sciencedirect.com	0,26	0,91
106	rd.springer.com	www.sciencedirect.com, link.springer.com	0,26	0,91
107	link.springer.com, rd.springer.com	www.sciencedirect.com	0,26	0,91
108	search.proquest.com, rd.springer.com	web.ebscohost.com	0,18	0,91
103	search.proquest.com, rd.springer.com	web.ebscohost.com, link.springer.com	0,18	0,91
104	search.proquest.com, link.springer.com, rd.springer.com	web.ebscohost.com	0,18	0,91
105	link.springer.com	www.sciencedirect.com	0,35	0,91

2014 – Reglas de Asociación

No.	Premisa	Conclusión	Soporte	Confianza
57	ebscohost.com, els-cdn.com	sciencedirect.com	0,29	1,00
58	els-cdn.com, proquest.com	sciencedirect.com	0,26	1,00
59	els-cdn.com, springer.com	sciencedirect.com	0,22	1,00
67	ebscohost.com, els-cdn.com, proquest.com	sciencedirect.com	0,21	1,00
61	ebscohost.com, ac.els-cdn.com	www.sciencedirect.com	0,19	1,00
60	els-cdn.com, wiley.com	sciencedirect.com	0,19	1,00
56	pubmed.gov	nih.gov	0,18	1,00
62	proquest.com, ac.els-cdn.com	www.sciencedirect.com	0,17	1,00
65	ac.els-cdn.com, link.springer.com	www.sciencedirect.com	0,17	1,00
63	ac.els-cdn.com, content.ebscohost.com	www.sciencedirect.com	0,17	1,00
66	ebscohost.com, els-cdn.com, www.sciencedirect.com	sciencedirect.com	0,17	1,00
64	ac.els-cdn.com, web.a.ebscohost.com	www.sciencedirect.com	0,16	1,00
68	sciencedirect.com, ebscohost.com, ac.els-cdn.com	www.sciencedirect.com	0,16	1,00
55	els-cdn.com, www.sciencedirect.com	sciencedirect.com	0,26	1,00
54	els-cdn.com, ac.els-cdn.com	sciencedirect.com	0,23	1,00
53	els-cdn.com, www.sciencedirect.com, ac.els-cdn.com	sciencedirect.com	0,23	1,00
52	media.proquest.com	search.proquest.com	0,22	1,00
51	ac.els-cdn.com	www.sciencedirect.com	0,38	1,00
50	ac.els-cdn.com, search.proquest.com	www.sciencedirect.com	0,19	1,00
49	els-cdn.com	sciencedirect.com	0,46	1,00
48	sciencedirect.com, ac.els-cdn.com	www.sciencedirect.com	0,26	1,00
47	els-cdn.com, ac.els-cdn.com	www.sciencedirect.com	0,23	1,00
46	sciencedirect.com, els-cdn.com, ac.els-cdn.com	www.sciencedirect.com	0,23	1,00
45	els-cdn.com, ac.els-cdn.com	sciencedirect.com, www.sciencedirect.com	0,23	1,00
44	nih.gov	pubmed.gov	0,18	0,99
43	sciencedirect.com, ebscohost.com, www.sciencedirect.com	els-cdn.com	0,17	0,93
42	sciencedirect.com, springer.com	els-cdn.com	0,22	0,92
41	www.sciencedirect.com, link.springer.com	ac.els-cdn.com	0,17	0,91
40	sciencedirect.com, wiley.com	els-cdn.com	0,19	0,91
39	sciencedirect.com, ebscohost.com, www.sciencedirect.com	ac.els-cdn.com	0,16	0,91
38	sciencedirect.com, ac.els-cdn.com	els-cdn.com	0,23	0,91
37	sciencedirect.com, www.sciencedirect.com, ac.els-cdn.com	els-cdn.com	0,23	0,91
36	sciencedirect.com, ac.els-cdn.com	els-cdn.com, www.sciencedirect.com	0,23	0,91
35	www.sciencedirect.com, web.a.ebscohost.com	ac.els-cdn.com	0,16	0,90
34	www.sciencedirect.com, content.ebscohost.com	ac.els-cdn.com	0,17	0,90
33	www.sciencedirect.com, proquest.com	ac.els-cdn.com	0,17	0,90
32	els-cdn.com, www.sciencedirect.com	ac.els-cdn.com	0,23	0,90
31	sciencedirect.com, els-cdn.com, www.sciencedirect.com	ac.els-cdn.com	0,23	0,90
30	els-cdn.com, www.sciencedirect.com	sciencedirect.com, ac.els-cdn.com	0,23	0,90

29	sciencedirect.com, ebscohost.com, proquest.com	els-cdn.com	0,21	0,90
28	sciencedirect.com, www.sciencedirect.com	els-cdn.com	0,26	0,90
27	www.sciencedirect.com, search.proquest.com	ac.els-cdn.com	0,19	0,90
26	ebscohost.com, springer.com	sciencedirect.com	0,16	0,90
25	sciencedirect.com, www.sciencedirect.com	ac.els-cdn.com	0,26	0,89
24	sciencedirect.com, proquest.com	els-cdn.com	0,26	0,89
23	ebscohost.com, www.sciencedirect.com	ac.els-cdn.com	0,19	0,89
22	sciencedirect.com, ebscohost.com	els-cdn.com	0,29	0,88
21	wiley.com	sciencedirect.com	0,20	0,88
20	ovid.com	sciencedirect.com	0,16	0,88
19	springer.com	sciencedirect.com	0,24	0,88
18	web.a.ebscohost.com, web.b.ebscohost.com	content.ebscohost.com	0,17	0,87
17	sciencedirect.com	els-cdn.com	0,46	0,87
16	www.sciencedirect.com, proquest.com	sciencedirect.com	0,17	0,86
15	link.springer.com	www.sciencedirect.com	0,19	0,86
14	www.sciencedirect.com	ac.els-cdn.com	0,38	0,86
11	ebscohost.com, ac.els-cdn.com	sciencedirect.com	0,16	0,85
12	ebscohost.com, ac.els-cdn.com	sciencedirect.com, www.sciencedirect.com	0,16	0,85
13	ebscohost.com, www.sciencedirect.com, ac.els-cdn.com	sciencedirect.com	0,16	0,85
10	ebscohost.com, www.sciencedirect.com	sciencedirect.com	0,18	0,84
7	els-cdn.com, proquest.com	ebscohost.com	0,21	0,81
8	els-cdn.com, proquest.com	sciencedirect.com, ebscohost.com	0,21	0,81
9	sciencedirect.com, els-cdn.com, proquest.com	ebscohost.com	0,21	0,81
6	sciencedirect.com, www.sciencedirect.com	els-cdn.com, ac.els-cdn.com	0,23	0,81
4	springer.com	els-cdn.com	0,22	0,81
5	springer.com	sciencedirect.com, els-cdn.com	0,22	0,81
3	sciencedirect.com, proquest.com	ebscohost.com	0,24	0,81
1	wiley.com	els-cdn.com	0,19	0,80
2	wiley.com	sciencedirect.com, els-cdn.com	0,19	0,80