



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

MEJORAMIENTO DE LA CONVERSIÓN DE ANUNCIOS PUBLICITARIOS EN UN
MEDIO DIGITAL

PROYECTO DE GRADO PARA OPTAR AL GRADO DE MAGÍSTER EN
INGENIERÍA DE NEGOCIOS CON TECNOLOGÍAS DE INFORMACIÓN

ANDRÉS PATRICIO SILVA MARAMBIO

PROFESOR GUÍA:
ÁNGEL JIMÉNEZ MOLINA

MIEMBROS DE LA COMISIÓN:
EZEQUIEL MUÑOZ KRSULOVIC
TODD PEZZUTI LLOYD

SANTIAGO DE CHILE
2018

Resumen Ejecutivo

El servicio de publicidad digital de Portalfruticola.com representa la principal fuente de ingresos de Yentzen Group. Considerando la tendencia mundial de la publicidad digital y los incipientes descontentos de clientes al inicio de este trabajo, se prevé una potencial fuga y dificultad de captura de clientes en un futuro a mediano-largo plazo.

Las razones del descontento se enmarcan principalmente en la baja efectividad de las campañas publicitarias (promedio de conversión igual a 0,24 % mensual/campaña para el año 2015) y un alto costo por clic equivalente (sobre los \$10.200 aproximadamente para las mejores campañas del 2015).

El proyecto consistió en realizar un experimento cuyo objetivo es identificar ya los factores, su combinación y niveles, que aumentan la conversión de las campañas publicitarias en Portalfruticola.com. Una de las variables es la recomendación contextual, por lo que se construirá un sistema de recomendación semántico para ello. El experimento tuvo lugar en un nuevo espacio publicitario dada restricción de que no se puede alterar la estructura y orden de los espacios publicitarios actuales.

El diseño experimental fue de tipo factorial 2^3 evaluando tres factores: formato (imagen vs texto), relevancia semántica (baja vs alta) y etiquetado (con vs sin etiquetado). Los datos analizados correspondieron al periodo entre septiembre a diciembre 2016 y fueron analizados en dos bloques de corridas debido a diferencias de normalidad de los datos.

Las combinación óptima de factores es “Anuncios de texto semánticamente relevantes sin etiquetado” y “Anuncios de texto semánticamente no relevantes sin etiquetado”. El etiquetado no produjo diferencias significativas a diferencia del formato pero se recomienda eliminarlo dada la carga visual del sitio web. El formato Texto se vio potenciado cuando se combinaba con el factor Relevancia Alta logrando CTR promedio de 9,1 % y 7,5 % para cada grupo.

Adicionalmente, el nuevo espacio publicitario demostró contribuir significativamente a la producción de clics, ya que generó un equivalente al 29,3 % de los clics totales generados por el resto de las 19 posiciones publicitarias del sitio web durante el mismo periodo.

Finalmente, se puede afirmar que este nuevo espacio publicitario es competitivo en el mercado, dado que se logró superar la conversión promedio mundial de Google AdSense de 0,35 % y de 7 % para Google Adwords, los que son potentes argumentos de venta. Lamentablemente, este trabajo no produjo un cambio en el comportamiento de la empresa.

Tabla de Contenido

1. Introducción y contexto	1
1.1. Antecedentes de la industria	1
1.2. Descripción general de la empresa	2
1.3. Problema u oportunidad identificada	3
1.4. Objetivos y resultados esperados del Proyecto	4
1.4.1. Objetivo general	4
1.4.2. Objetivos específicos	4
1.4.3. Resultados esperados	5
1.5. Alcance	6
1.6. Riesgos potenciales	6
2. Marco teórico	7
2.1. Metodología de Ingeniería de Negocios	7
2.2. Lógica de negocios	9
2.2.1. Un problema de asignación	9
2.2.2. Factores que gobiernan el éxito de un anuncio	11
2.2.3. Evidencias de soluciones existentes	12
3. Planteamiento Estratégico y Modelo de Negocio	14
3.1. Posicionamiento Estratégico	14
3.2. Balanced Scorecard	15
3.3. Modelo de Negocios	16
4. Análisis Situación Actual	19
4.1. Arquitectura de Procesos	19
4.2. Modelamiento detallado de procesos	21
4.2.1. Modelamiento IDEF0	21
4.3. Diagnóstico de la situación actual	26
4.4. Cuantificación del problema u oportunidad	27
5. Propuesta de Diseño de Procesos	28
5.1. Direcciones de cambio y alcance	28
5.2. Arquitectura de procesos To be	29
5.3. Diseño de procesos detallados To be	30
5.3.1. Modelamiento IDEF0	30
5.3.2. Modelamiento BPMN	30
5.4. Diseño de lógica de negocios	31

6. Propuesta de apoyo tecnológico	39
6.1. Especificación de requerimientos	39
6.1.1. Requerimientos funcionales	39
6.1.2. Requerimientos no funcionales	39
6.2. Arquitectura tecnológica	41
6.3. Diseño de la aplicación	41
6.3.1. Casos de Uso	41
6.3.2. Diagramas de Secuencia	41
6.3.3. Diagramas de Clases	42
7. Evaluación del Proyecto	50
7.1. Definición de beneficios y costos	50
7.2. Flujo de caja	53
7.3. Análisis de Sensibilidad	56
8. Resultados	57
8.1. Extracción, transformación y limpieza de datos	57
8.2. Análisis estadístico	63
8.2.1. Análisis corridas 1 a 4	64
8.2.2. Análisis corridas 5 a 8	66
8.2.3. Análisis para el factor “Etiquetado”	67
8.2.4. Clics totales	68
9. Discusión	69
10.Trabajo futuro	72
11.Conclusiones	74
12.Bibliografía	75

Índice de Tablas

1.1. Detalle de los principales riesgos identificados y sus medidas de mitigación de corto plazo.	6
5.1. Caracterización de las variables de respuesta.	33
5.2. Niveles de los factores de diseño o variables de control para el experimento. .	35
5.3. Método de medición de los factores de diseño o variables de control.	35
5.4. Corridas del experimento	36
7.1. Parámetros y supuestos utilizados en la evaluación financiera del proyecto. .	51
7.2. Cálculo del precio de venta por CPM.	51
7.3. Estimación de ingresos y costos incrementales atribuibles al proyecto.	52
7.4. Valores de las variables para calcular la tasa de descuento a través del método CAPM.	53
7.5. Flujo de Caja Privado. Todos los valores son en pesos chilenos.	55
7.6. Sensibilización del supuesto de renovaciones de contrato atribuibles al proyecto.	56
9.1. CTR promedio de anuncios de texto con diferentes relevancias	71

Índice de Ilustraciones

1.1. Inversión por medio en 2015 en Chile.	2
1.2. Evolución del indicador CTR para todos los países y todas las industrias (Google, 2016)	3
2.1. Metodología de Ingeniería de Negocios (Barros, 2009)	8
2.2. Estructura de la página evaluada en el experimento de “seguimiento de vista” de Drèze y Hussherr (2003).	11
2.3. Motor de Publicidad Contextual desarrollado por Lee, Ha, Jung, y Lee (2013).	13
3.1. Posicionamiento Estratégico Actual y Futuro de Yentzen Media.	14
3.2. Balanced Scorecard de Yentzen Media	16
3.3. Modelo de Negocios propuesto para Yentzen Media	17
4.1. Nivel A-0 de Yentzen Media	20
4.2. Macroprocesos de Yentzen Media	22
4.3. Detalle de la cadena de valor “Servicio de publicidad digital”	23
4.4. Detalle del proceso de “Administración de la relación con el cliente” de la cadena de valor “Servicio de publicidad digital”	24
4.5. Detalle del proceso de “Marketing y análisis de mercado” de la cadena de valor “Servicio de publicidad digital”	25
4.6. Detalle del proceso de “Producción, publicación y despliegue” de la cadena de valor “Servicio de publicidad digital”.	26
5.1. Posibilidades de Dirección de Cambio.	28
5.2. Arquitectura propuesta para Yentzen Media. El color naranja se destacan los flujos y procesos que serán agregados o modificados.	30
5.3. Estructura interna propuesta para el proceso de “Marketing y análisis de mercado” de Yentzen Media. Los nuevos flujos se encuentran destacados con color naranja, los cuales provienen del servicio compartido “Análisis del comportamiento del cliente y lector” (Figura 5.2).	31
5.4. Estructura interna propuesta para el proceso de “Producción, publicación y despliegue” de la cadena de valor “Servicio de publicidad digital” de Yentzen Media. En naranja se destaca lo que será influenciado por el proyecto.	32
5.5. Diagrama de causa y efecto identificados que influyen en la variable de respuesta en el experimento.	34
5.6. Proceso de recomendación de anuncios para el experimento.	37

6.1. Diagrama de actores de la arquitectura tecnológica.	43
6.2. Diagrama de casos de uso.	44
6.3. Diagrama de Secuencia del caso de uso “Obtener recomendaciones”	45
6.4. Diagrama de Secuencia del caso de uso “Actualizar colección posts”	46
6.5. Diagrama de Secuencia del caso de uso “Actualizar perfiles”	47
6.6. Diagrama de Secuencia del caso de uso “Insertar anuncios”	48
6.7. Diagrama de Clases	49
8.1. Distribución de variables del conjunto de datos.	60
8.2. Distribución de variables del conjunto de datos.	61

Capítulo 1

Introducción y contexto

1.1. Antecedentes de la industria

Antecedentes de la industria hortofrutícola de Chile

La agricultura es una antigua e importante actividad que está ligada al negocio de los alimentos, lo que es mucho más amplio que sólo su producción. El negocio de la agricultura también incluye a diversos profesionales que trabajan en diferentes procesos de la cadena productiva, aportando productos como también servicios al rubro. Según Trademap.com (2015) a nivel mundial, el 2014 se exportaron \$104.129.224 miles de dólares, donde los países más relevantes fueron Estados Unidos, España, Holanda y Chile.

Chile un jugador importante en la industria, de hecho es el exportador número 1 de fruta fresca del hemisferio sur, y dentro de su economía interna, representa el tercer sector más importante, con un total estimado de 7.800 agricultores de fruta fresca, y un total de 310.000 hectáreas cultivadas a lo largo de este país (Fruits From Chile, 2014). En detalle, Chile exporta más de 30 especies diferentes a más de 100 países diferentes en el mundo (Fruits From Chile, 2014).

Antecedentes de la industria de la publicidad digital

Según el estudio de la Asociación Chilena de Agencias de Publicidad (ACHAP) (2015), tanto en Chile como en el mundo, el principal medio de inversión publicitaria es la televisión, con un 37,4% y 37,2% respectivamente, seguido por los medios digitales 13,7% y 29,4% respectivamente.

Según el mismo estudio, hay una clara tendencia al aumento de la inversión en medios online, de hecho hubo un aumento del 29,3% de las inversiones digitales con respecto al 2014 en Chile. Por lo tanto, Chile no está ajeno a esta tendencia mundial, por lo que la esta industria está creciendo.

En términos de magnitudes monetarias, el 2015 la inversión online alcanzó US \$161 millones, superando por amplia mayoría a otros medios (Figura 1.1). Esto representa una inversión publicitaria per cápita igual a 65,5 US\$, muy inferior aún a los 400 US\$ per cápita de Estados Unidos.

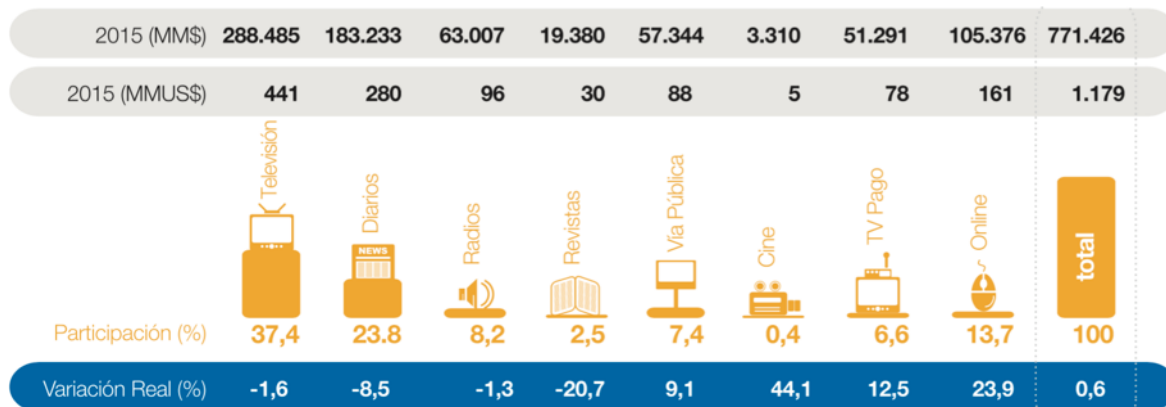


Figura 1.1: Inversión por medio en 2015 en Chile.

Un concepto importante en publicidad digital es el indicador de efectividad o desempeño de una campaña publicitaria, conocido como *Click Through Rate* (CTR). Este se calcula dividiendo la cantidad de clics totales registrados por un anuncio sobre la cantidad de veces que este fue cargado en el sitio web (impresiones) en un tiempo determinado (Fórmula 1.1). Este indicador también es conocido como tasa de conversión.

$$\text{Click Through Rate (CTR)} = \frac{\text{Clics}}{\text{Impresiones}} \tag{1.1}$$

Utilizando la herramienta “Display benchmarks” de Google, se obtiene la gráfica de la Figura 1.2.

El promedio para el segundo trimestre del 2016 es igual a 0,35 % para el tipo de anuncios “banners”, mientras que la tasa de conversión para formatos “Rich Media”, como videos, es de 2,74 % (Google, 2016). Este es formato que registra mejores tasas de conversión.

1.2. Descripción general de la empresa

Yentzen Group fue fundada en el año 2006 y se divide operacionalmente en dos áreas: Yentzen Consulting y Yentzen Media. La primera realiza consultorías en marketing y estudios de imagen y posicionamiento de marca en el sector hortofrutícola internacional. Por otro lado, Yentzen Media es un conjunto de sitios web dedicados a comunicar el acontecer de la industria hortofrutícola internacional, éstos sitios web son: Portalfruticola.com, Freshfruitportal.com y Chinafruitportal.com.

Yentzen Media tiene dos actividades principales: (1) la generación de noticias, y (2) el servicio de publicidad digital cuyo objetivo es monetizar el trabajo de la primera actividad.

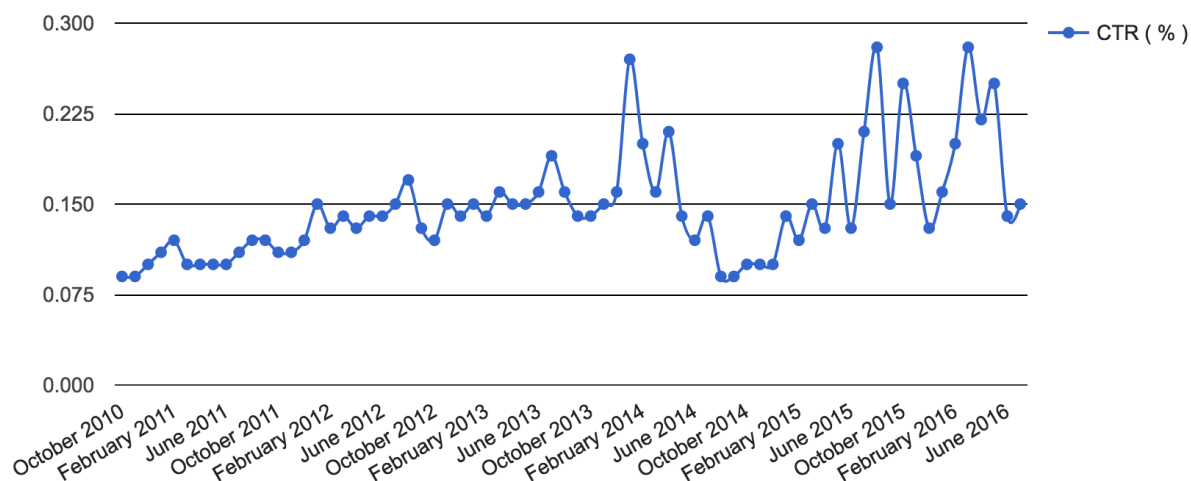


Figura 1.2: Evolución del indicador CTR para todos los países y todas las industrias (Google, 2016)

El presente proyecto se enmarcará en Portalfruticola.com, para luego extenderse al resto de lo sitios administrados por Yentzen Media.

1.3. Problema u oportunidad identificada

El principal ingreso de la Yentzen Group proviene de la contratación de campañas publicitarias, servicio operado por su subdivisión llamada Yentzen Media. Dichas campañas presentan una baja tasa de conversión promedio (0,24 % mensual/campaña para el año 2015) y un alto costo por clic equivalente (sobre los \$10.200 aproximadamente para las mejores campañas del 2015), lo que está generando quejas y disconformidad en los clientes. Se estima que este fenómeno afectará la renovación de contratos porque se relaciona estrechamente con la efectividad de la campaña. Por lo tanto, se quiere evitar, o al menos mitigar, la situación de fuga incipiente actual, y potencialmente severa, en el futuro a través del mejoramiento del desempeño de las campañas publicitarias.

Mejorar el desempeño de las campañas a niveles sobre el promedio de competidores también ayudará a mitigar las potenciales fugas de inversión hacia otras redes publicitarias, como Google AdSense o Google Adwords. En términos prácticos, si el desempeño de las campañas es más eficiente en otras redes publicitarias, con costos de inversión más bajos, la decisión lógica de los clientes sería cambiarse de plataforma publicitaria. Afortunadamente, los clientes en esta industria aún son adversos al uso de este tipo de plataformas, y aún pesa más la reputación de los sitios webs tradicionales o la amistad/confianza entre dueños. Pero este escenario puede cambiar en los próximos años, cuando un segmento más joven ascienda a cargos de responsabilidad y realice éstos cálculos (hechos que ya comienzan a vislumbrarse).

Por ello, la empresa debe prepararse para dicho momento mejorando el desempeño de las campañas a niveles muy superiores a los de las redes publicitarias tradicionales (0,35 % como

se menciona en la Página 2). Si se logra esto, el Equipo de Ventas podría respaldar el alto precio de venta con mejores rendimientos a un público de nicho.

Se estima que el bajo desempeño de las campañas publicitarias se debe principalmente a dos factores: primero, los lectores se enfrentan a una gran carga visual, ya que se despliega en promedio al menos 7 anuncios por página, y muchos de ellos quedan fuera de la zona visible inmediata. Es probable que los anuncios de interés no sean percibidos fácilmente, tanto por la carga visual o por que quedan más abajo en la página.

El segundo factor es que el sitio no posee publicidad contextual, ya que las opciones de segmentación de los anuncios publicitarios considera únicamente el país de visitante, posición del banner en la página y si esta posición es en la portada o en cualquier otra página. La literatura, y la tendencia en publicidad digital, indican que la publicidad contextual tiene mejores tasas de conversión, así lo indica incluso Google (Hof, 2013), además de señalar que el clic será la única métrica utilizada para evaluar el rendimiento de los anuncios. Esta es una oportunidad para la empresa aumente su cuota de mercado, ya que los competidores tampoco no poseen este tipo de segmentación.

La problemática no es sólo saber qué anuncios desplegar, sino también cómo presentarlo. Por lo que el problema de asignación se extiende también a un problema de “exploración o caracterización de factores” que influyen en la interacción del usuario (lector) con los anuncios publicitarios dentro del ecosistema digital de Portalfruticola.com. La necesidad de estudiar factores e interacciones elementales que influyen en el principal ingreso de la empresa, es porque nunca se ha hecho, y por decisión de la Dirección, no se puede modificar el layout de despliegue de los anuncios que actualmente se maneja.

1.4. Objetivos y resultados esperados del Proyecto

1.4.1. Objetivo general

Determinar los factores, su combinación y niveles, que aumentan la conversión de las campañas publicitarias en Portalfruticola.com en un nuevo espacio publicitario durante el 2016.

1.4.2. Objetivos específicos

Se plantean los siguientes objetivos específicos:

1. Diseñar un experimento que permita seleccionar los factores, su combinación y niveles, que aumentan la conversión de las campañas publicitarias durante las primeras dos semanas iniciado el proyecto.
2. Desarrollar un sistema de recomendación para seleccionar los anuncios a desplegar de

acuerdo a los intereses del lector y el contenido que se encuentra leyendo antes del mes de agosto 2016.

3. Realizar el experimento en Portalfruticola.com durante al menos 1 mes para recolectar datos de las recomendaciones entregadas y la interacción de los lectores con ellas.
4. Realizar un análisis estadístico para seleccionar los factores, su combinación y nivel, que se utilizarán definitivamente, antes del mes octubre 2016.
5. Modificar el sistema de recomendación de acuerdo a los hallazgos encontrados antes de diciembre 2016.

Considerando los objetivos específicos, se enuncian las siguiente interrogantes de investigación:

1. ¿Los anuncios de texto presentarían un mejor desempeño que los anuncios de imagen?
2. Dado que los anuncios estarán etiquetados como patrocinado, ¿habría un cambio de comportamiento después de un número determinado de sesiones del usuario? ¿habría un incremento en los banners del costado derecho?
3. ¿Los anuncios agrupados y presentados en categorías tendrían mejor desempeño que si fueran presentados como una lista simple?
4. Considerando que se está en un dominio específico del conocimiento ¿Es más relevante la cercanía semántica del anuncio con el par (Contenido,Perfil) que la calidad del llamado del mismo?
5. ¿Existe alguna diferencia de comportamiento en la interacción con los anuncios publicitarios de la comunidad de Portalfruticola.com con los comportamientos descritos por otros autores en otros campos?

Finalmente, se establecen las siguientes hipótesis:

1. Los anuncios que presenten una cercanía semántica alta entre el par (Contenido,Perfil) tendrán una mayor conversión que los que no.
2. Los anuncios de texto presentarán una mayor tasa de conversión que los de imagen.
3. Los anuncios que se presenten etiquetados tendrán una mayor conversión que los que no, independiente de su relevancia o formato.
4. Este nuevo espacio publicitario será más efectivo en términos de conversión que el resto de las posiciones dado su carácter dinámico.

1.4.3. Resultados esperados

1. Incrementar la conversión de la campañas para que iguallen o superen el 1%.

2. Aumentar la renovación de contratos en un 3% al presentar campañas más efectivas a los clientes.
3. Se espera que este nuevo espacio publicitario pueda soportar las nuevas lógicas de advertising web como la venta por CPM y CPC en el futuro cercano.

Dado el marco de tiempo en el que se desarrolla este proyecto, es probable que los últimos dos resultados esperados no puedan ser incluidos en los resultados de la presente tesis.

1.5. Alcance

Se creará un nuevo espacio publicitario dada la restricción de negocio de que no se puede alterar la estructura y orden de los espacios publicitarios actuales. Este espacio se ubicará al final de cada noticia, cuyo título será “También podría interesarte”.

El experimento y la implementación del sistema de recomendación se realizará únicamente en Portalfruticola.com para efectos de este trabajo.

1.6. Riesgos potenciales

Los riesgos potenciales identificados se presentan en la Tabla 1.1.

Tabla 1.1: Detalle de los principales riesgos identificados y sus medidas de mitigación de corto plazo.

Tipo de riesgo y nivel de riesgo	Descripción	Medida de mitigación de corto plazo
Tecnológico (Medio)	El proyecto requiere individualizar el tráfico a nivel de usuario, pero el sitio no tiene lógicas para ello, sólo se cuenta con Google Analytics con la configuración por defecto.	Se investigarán y habilitarán lógicas para lograr individualizar el tráfico.
Económico (Medio)	Hay incertidumbre respecto a los proyectos estratégicos que se priorizarán para la asignación de fondos.	Se acordó montar un prototipo en el sitio para evaluar los potenciales resultados, beneficios y costos.

Capítulo 2

Marco teórico

2.1. Metodología de Ingeniería de Negocios

La Metodología de Ingeniería de Negocios propuesta por Barros (2009) se resume en la Figura 2.1, que básicamente consiste en realizar una revisión exhaustiva desde la empresa desde varias puntos de vista. Esta revisión habilita un entendimiento ingenieril del negocio permitiendo idear y desarrollar proyectos que se alineen con la organización. Hay que considerar que esta metodología es de tipo “Top-Down”.

Considerar que la metodología se aplicó haciendo una instanciación de la realidad actual y se comparó con la situación futura deseada. Esto determina las brechas entre la estrategia y las oportunidades de mercado y organización del negocio. A través del análisis de éstas brechas que es que el presente proyecto fue propuesto.

Adicionalmente, la implementación de esta metodología se pueden combinar otras metodologías específicas para resolver cada uno de los pasos propuestos, los cuales se detallarán a continuación:

1. **Planteamiento Estratégico:** Busca comprender y determinar el cómo la empresa se plantea en el mercado y a lo que aspira. Para ello primero se recurre a la revisión de la visión y misión, descripción de líneas de negocio y *Balanced Scorecard*. Con ello se conceptualiza el estado actual y deseado de la empresa utilizando el modelo Delta Hax (Hax, 2009). El output lógico de la estrategia de la empresa deben ser planes a implementar.
2. **Definición del Modelo de Negocio:** Como segundo paso, se define el modelo de negocio para comprender la propuesta de valor, los segmentos de clientes y el funcionamiento conceptual de las piezas claves que habilitan el negocio. Para ello se conceptualizará utilizando el Modelo Canvas (Osterwalder y Pigneur, 2010).
3. **Diseño Arquitectura de Procesos:** Como tercer paso, se instancian la arquitectura de procesos de la empresa utilizando la metodología propuestas por Barros (2009),

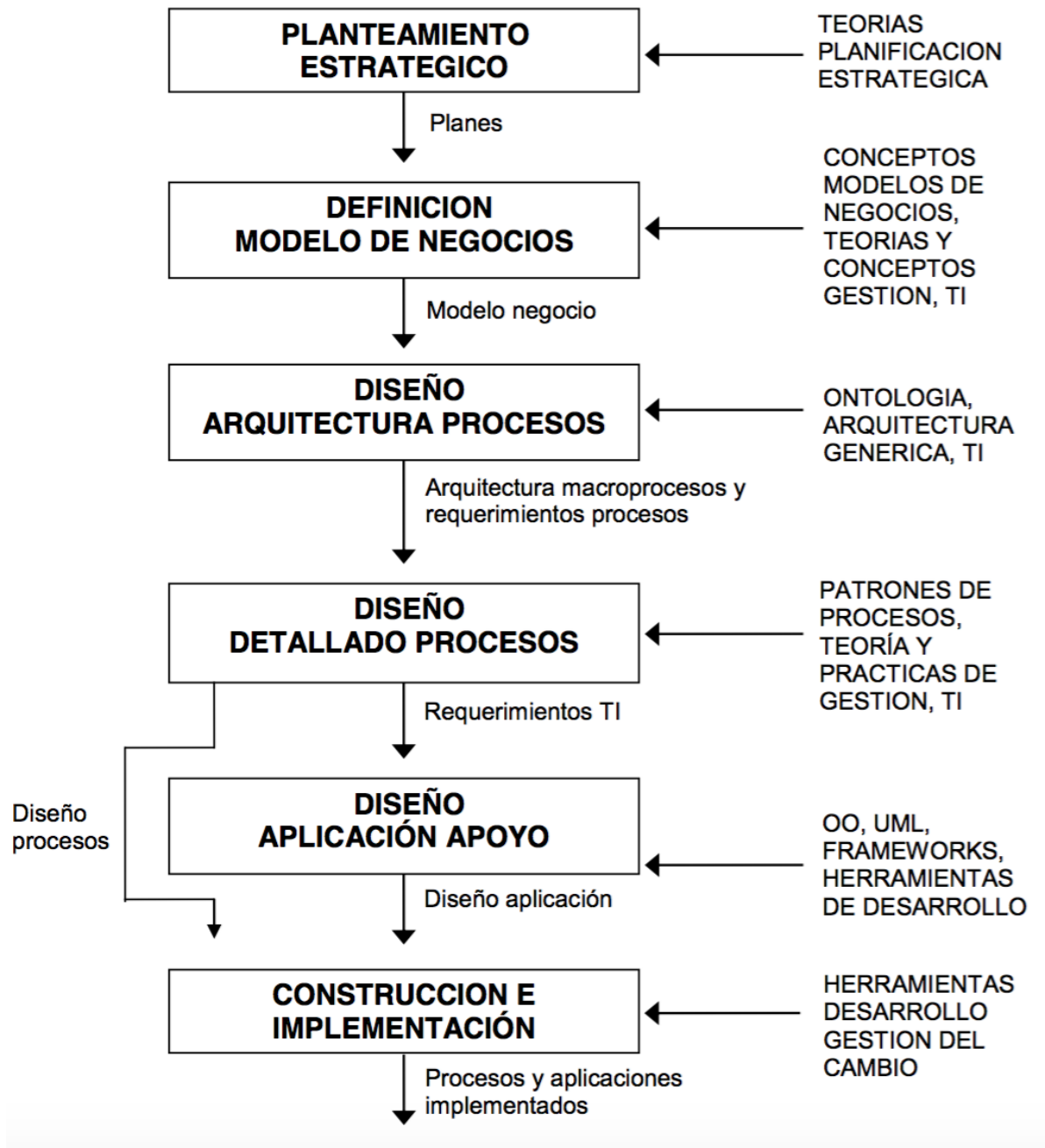


Figura 2.1: Metodología de Ingeniería de Negocios (Barros, 2009)

la que contiene los patrones de arquitectura de procesos completos. De esta manera, se puede determinar cuáles flujos o procesos no están completos en la actualidad. Adicionalmente, se debe considerar que un proyecto puede considerar modificar esta arquitectura. El diagramar esto, permite comprender las diferentes interacciones que deberán instalarse en la empresa. Para ello se recurre a la notación IDEF0 (Mayer, 1992).

4. **Diseño Detallado de Procesos:** El cuarto paso es muy similar al tercero, sólo que los macroprocesos detallados anteriormente, mirados por dentro, lo que aumenta el nivel de comprensión que se busca sobre el funcionamiento de la empresa.
5. **Diseño Aplicación de Apoyo:** En el quinto paso se diseñan las aplicación de apoyo para poder operar el negocio bajo los indicadores deseados. Para ello se utilizan el lenguaje de UML.
6. **Construcción e implementación:** Como último paso, se procede a la implementación de las soluciones diseñadas.

2.2. Lógica de negocios

2.2.1. Un problema de asignación

La tendencia en la publicidad digital es el cobro por resultados concretos de la publicidad. Históricamente se ha intentado poder establecer la efectividad y resultados de un anuncio publicitaria en diferentes medios, y es en el mundo digital donde esto ha avanzado fuertemente dada las posibilidades de seguimiento que hoy ofrece la web. Esto ha dado paso natural a que los anunciantes cuestionen e intenten determinar si su inversión publicitaria fue realmente efectiva. Pese a que esta tendencia en el mundo publicitario aún no está completamente internalizada y desarrollada como otras industria, el hecho es inminente.

Dado lo anterior es que las redes publicitarias comenzaron a desarrollar los modelos de cobros por resultados, porque era posible saber cuándo se cumplía dicho objetivo. Pese a que no es un término estricto, a esto se la denominado en la web como “el marketing de resultados” o “publicidad en base a objetivos”.

Lo anterior implica que entonces debe haber una asignación óptima en tiempo real entre el universo de anuncios disponibles y las impresiones generadas por los usuarios en las páginas web; entonces el problema que enfrentan las redes publicitarias, es un problema de asignación. Además no solamente la asignación debe seleccionar los anuncios utilizando un modelo que permita optimizar la rentabilidad y consecución de objetivos, sino que se debe seleccionar una determinada cantidad de Ads por impresión, ya que cada impresión tiene un inventario disponible determinado, el cuál varía de sitio en sitio.

Si es que las opciones de inventario disponibles fueran “muchas” y éstas no fueran superadas por el universo de Ads disponibles, no habría problema de optimización. Pero ¿cuánto es

“mucho”?, y ¿realmente es beneficioso desplegar “muchos” anuncios por impresión?

Pese a que no hay un estándar que defina cuándo una página tiene “muchos anuncios”, y las opiniones sobre esto varían en Internet, Google AdSense entrega lineamientos de buenas prácticas e indica que no se debieran desplegar más de 3 anuncios por impresión de su red (Google AdSense, 2015). Sin duda Google es un actor relevante que marca la pauta en el mundo de la publicidad digital, pero no significa que se puedan incorporar más unidades por impresión, pero también hay que considerar que Google considera sus buenas prácticas para priorizar también los resultados en buscadores.

Experimentos usando la neurociencia han realizado avances en esta área, ya que buscan comprender como los usuarios responden al uso de las páginas, por lo que poco a poco habrán en la literatura científica hallazgos sobre los lineamientos de diseño de páginas web para conseguir ciertos objetivos, lineamientos de diseño basado en resultados de experimentos cuantitativos.

Drèze y Hussherr (2003) realizaron una serie de experimentos en el que determinaron que un anuncio tiene una probabilidad 50 % independiente de la experiencia de navegación, género y edad de los navegantes. Pero esto válido para las condiciones que ellos probaron que fueron un anuncio ubicado en la parte superior como se muestra en la Figura 2.2. Evidentemente un usuario no puede interactuar con un anuncio si no lo ve, y esta probabilidad calculada por éstos autores, podría ser aún menor si el anuncio se encuentra alejado del contenido que el usuario está buscando en la página web y/o hay más anuncios por página.

Los mismos autores también descubrieron mediante experimentos en laboratorio, que las personas tienden a entender la estructura de la página y aprender a obviar los anuncios. Esto varía según la experiencia de navegación del usuario.

Adicionalmente, Goldstein, McAfee, y Suri (2013) realizaron un estudio para determinar el costo de los anuncios molestos, que finalmente producen un abandono del sitio, lo que repercute directamente en los ingresos del medio. Es por ello que no se debe abusar de la repetición y de la sobrecarga del sitio con anuncios publicitarios.

Naturalmente, los pop-ups que interrumpen al usuario caen en la categoría de “anuncios molestos”, pero Courbet, Fourquet-Courbet, Kazan, y Intartaglia (2014) determinó que este formato es el que mejor resultado tiene en la recordación de marca. Así mismo concluye Acquisti y Spiekermann (2011), que determinaron que la publicidad agresiva, especialmente la que interrumpe la actividad del usuario, puede aumentar la conciencia sobre la marca pero disminuye la voluntad a pagar/comprar algo de esa marca.

Por lo tanto, en base a la evidencia, cada medio deberá crear sus propias políticas sobre los anuncios pagados porque hay un costo de oportunidad con cada decisión.



Figura 2.2: Estructura de la página evaluada en el experimento de “seguimiento de vista” de Drèze y Hussherr (2003).

2.2.2. Factores que gobiernan el éxito de un anuncio

El éxito de un anuncio depende exclusivamente del objetivo del mismo. Basándose en las secciones anteriores, podríamos categorizar básicamente el objetivo de un anuncio en dos principales categorías: (1) basados en la acción, y (2) no basados en la acción (impresiones). Los primeros son los más atractivos para los clientes anunciantes y cuando la acción se cumple se le llama “conversión”.

El indicador Click Through Rate (CTR) se utiliza comúnmente para medir la efectividad del anuncio, el cual consiste básicamente en dividir la cantidad de clics recibidos sobre la cantidad de impresiones. Evidentemente este es un indicador válido para las campañas en base a objetivos, y los medios que ofrezca una tasa más alta de clics, serán percibidos como atractivos para la inversión por los anunciantes (considerando también otros factores por supuesto).

Pese a que el indicador CTR es el más utilizado actualmente para medir la efectividad de una campaña Drèze y Hussherr (2003) lo critican. Ellos plantean que esta es una medida a corto plazo, y no mide los demás beneficios a largo plazo en el recuerdo, reconocimiento y conciencia de la marca. Lamentablemente no lograron plantear el cálculo de un indicador aplicable fuera del laboratorio.

Recordemos que los mismos autores determinaron que la probabilidad de que el anuncio sea visto por el usuario, una página con un solo anuncio en la parte superior (Figura 2.2), es del 50 %. La tentación inicial sería aumentar la cantidad de anuncios o aumentar su tamaño para aumentar la probabilidad de ser visto, y con ello, existiría la posibilidad de que el usuario interactúe con la publicidad. Sin embargo, Drèze y Hussherr (2003) concluyeron que el factor más importante para la conversión es el mensaje publicitario, no los factores estéticos como el tamaño, color, contraste, diseño y animación.

Sin embargo, en base a la evidencia, se plantea que tanto la selección, cantidad por impresión, su ubicación dentro de la página y la competencia con otros objetos web, si deben afectar la conversión. Esto será probado en otra investigación en la empresa en conjunto con la Universidad de Chile.

2.2.3. Evidencias de soluciones existentes

El problema de asignación de anuncios se ha resuelto a través de sistemas de recomendación basados principalmente en (1) etiquetas, (2) intereses del usuario y (3) semántica. Este es último ha resultado de alto interés en las investigaciones recientes dado que permite encontrar pares de anuncios y contenido de manera más exacta. A esto se le llama anuncios contextuales, cuyo objetivo es insertar anuncios comerciales en las páginas web con contenido afín para ofrecer una mejor experiencia de usuario, lograr más conversiones (Xu, Wu, Li, y Chen, 2014). Estudios han demostrado que la cantidad de clics de los anuncios puede ser aumentada si se despliegan en páginas con contenido relacionado (Chatterjee, Hoffman, y Novak, 2003; Yih, Goodman, y Carvalho, 2006).

La razón de por qué la semántica estaría teniendo mejores resultados que pares página-anuncio determinado sólo por coincidencia de etiquetas, es debido al *Bijjective problem*, lo que implica que se debe partir del supuesto de que una etiqueta equivale a un concepto, cuando en la realidad eso no ocurre así (Hoppe, Roxin, y Nicolle, 2014).

Para lograr una recomendación semántica, tanto el contenido de la página web como el contenido del anuncio deber ser clasificado semánticamente. Esto es interesante porque se relaciona directamente con el hallazgo de Drèze y Hussherr (2003) sobre que es el mensaje publicitario lo más importante en la conversión.

Los modelos en este campo suelen desarrollarse utilizando alguna base de datos semántica existente como Wikipedia (Z. Wu, Xu, Zhang, Dolog, y Lu, 2012; Xu y cols., 2014), Open Directory Project (Lee y cols., 2013) o YAGO Ontology (Hoppe y cols., 2014). Sin embargo, Hoppe y cols. (2014) enriqueció la base semántica existente mediante el modelamiento del conocimiento de expertos en los dominios que trabajó.

Por otro lado, el motor desarrollado por Lee y cols. (2013) se muestra en la Figura 2.3, el cuál es interesante para la presente tesis. Este modelo consiste en utilizar el Open Directory Project¹ para desarrollar una taxonomía de tópicos y un clasificador que categoriza las páginas

¹Open Directory Project es un proyecto colaborativo a nivel internacional en cual editores voluntarios clasifican páginas web, lo que permite entrar modelos de clasificación.

web y los anuncios. Luego se realiza un ranking del par página-anuncio para determinar los anuncios relevantes a la página.

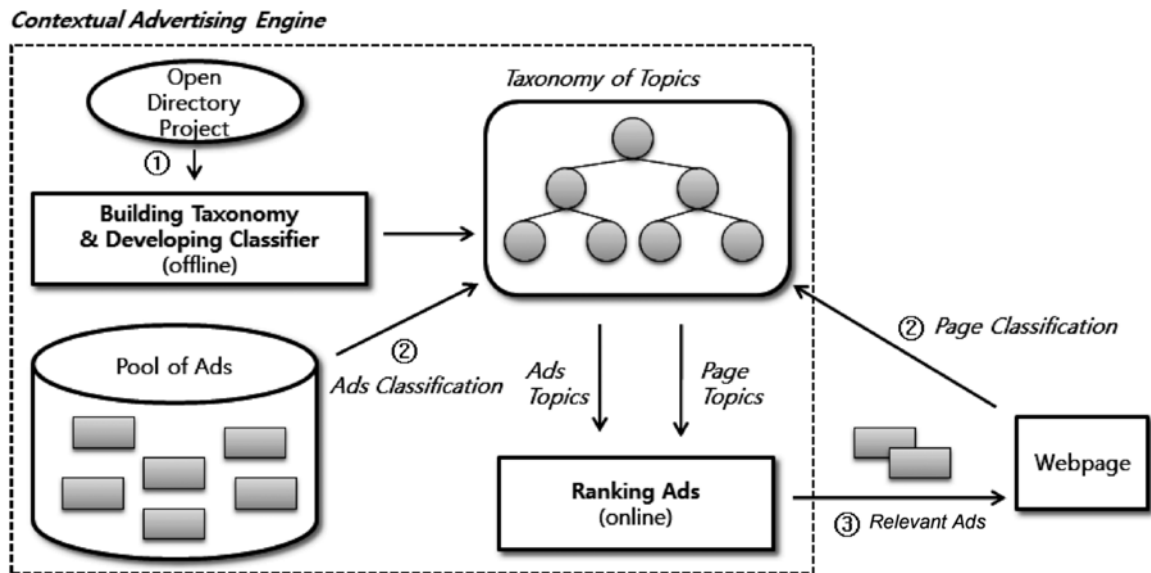


Figura 2.3: Motor de Publicidad Contextual desarrollado por Lee y cols. (2013).

Sin embargo, el modelo desarrollado por Lee y cols. (2013) no considera al usuario para determinar los anuncios relevantes. Se estima que para este trabajo de tesis será necesario considerar los intereses del lector explícito (declarados en su perfil) como también los intereses no explícitos que se pueden conseguir analizando su comportamiento web como lo realizó de Hoppe y cols. (2014) y X. Wu y cols. (2009).

El modelo ontológico desarrollado por Hoppe y cols. (2014) enriqueció la base de datos ontológica utilizada mediante el modelamiento semántico del conocimiento experto, lo que se estima que deberá aplicarse en este trabajo puesto que el dominio en el que desarrollará es muy específico y es muy probable que no sirvan modelo genéricos como los mencionados anteriormente.

Capítulo 3

Planteamiento Estratégico y Modelo de Negocio

3.1. Posicionamiento Estratégico

Utilizando la metodología de Hax (2009), Yentzen Media se posiciona actualmente como “Mejor Producto”, posicionamiento que desea reforzarse en el futuro (Figure 3.1). Esta definición implica que se trabajará en crear/agregar el mejor contenido, en consolidar la mejor comunidad agrícola digital y ofrecer el servicio de publicidad digital más eficaz y eficiente de la industria.

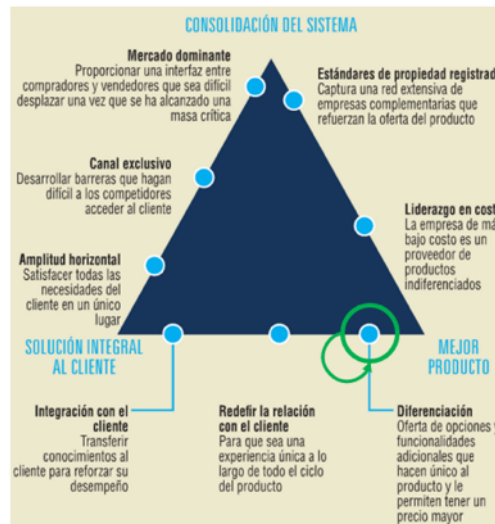


Figura 3.1: Posicionamiento Estratégico Actual y Futuro de Yentzen Media.

La generación de contenido se enfoca en la creación de contenido original, oportuno, objetivo y completo. La originalidad es un factor diferenciador en el área de las noticias, ya que ellas son de factura propia. Esto es debido a que se realiza investigación periodística y no una mera recopilación de comunicados de prensa o contenidos de otros medios.

Además cabe destacar que los sitios no son una simple traducción de las noticias a los otros idiomas, sino que cada equipo de periodistas genera su propio contenido para los lectores de los mercados que cubren.

Para el caso del servicio de publicidad digital, este también se posiciona actualmente como “Mejor Producto”, ya se ofrece un servicio de alto alcance, con segmentación por países y con informes de resultados, características que no tiene la competencia. Pese a esta diferenciación, clientes se pierden por precio, situación que se abordará más adelante.

3.2. Balanced Scorecard

Actualmente Yentzen Media no posee un Balanced Scorecard propuesto por la Dirección, por lo tanto, lo que aquí se presenta es una propuesta para Yentzen Media.

Como medio de comunicación, los objetivos estratégicos de Yentzen Media tienen relación más con el posicionamiento del servicio que con el beneficio económico, ya que presenta una sinergia entre las líneas de negocio: el éxito de que los lectores frecuenten la página, hace más atractivo para empresas y personas querer estar presente en el mismo, lo que promueve las ventas del servicio de publicidad digital. Por lo tanto, el posicionamiento del sitio web es fundamental para las ventas.

Por lo anterior, se definen los siguientes objetivos estratégicos principales:

1. Ser la plataforma de actualidad agrícola líder a nivel mundial para el 2018.
2. Ser la plataforma de actualidad agrícola líder en China para el 2018.
3. Mejorar la efectividad de las campañas de publicidad de digital en un 10 % para finales del 2016.
4. Crear nuevas alternativas de incorporación de contenido pagado (publicidad digital) para finales de 2016.

Considerando los objetivos anteriores, en la Figura 3.2 se presenta el Balanced Scorecard borrador propuesto para Yentzen Media, el que se encuentra en directa relación con el posicionamiento estratégico futuro deseado. El proyecto de innovación influyen los objetivos destacados.

El presente trabajo impacta sobre los objetivos destacados en la Figura 3.2, los cuáles son:

- Mejorar la efectividad de las campañas a un 1,5 % para el 2017.
- Implementar lógicas de despliegue de contenido y pricing adecuados para cada segmento de cliente.
- Mejorar el conocimiento sobre los lectores y clientes.

A su vez, éstos objetivos inciden en el principal que es “Crear nuevas alternativas de

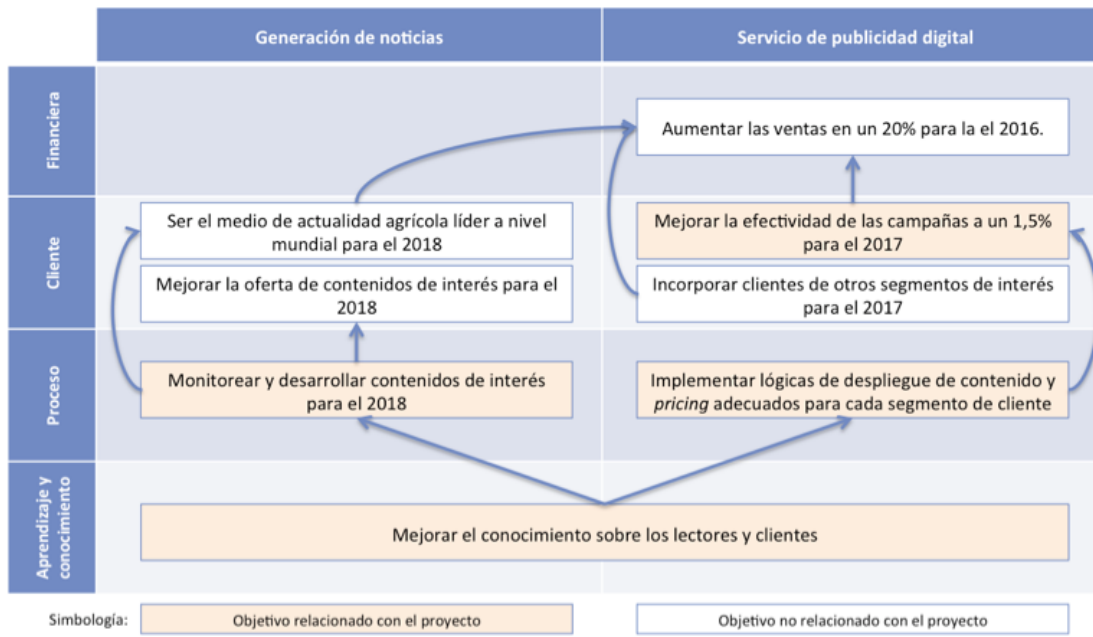


Figura 3.2: Balanced Scorecard de Yentzen Media

incorporación de contenido pagado (publicidad digital) para finales de 2016” mencionado anteriormente.

3.3. Modelo de Negocios

El modelo de negocios no se encuentra formalmente en la organización, y aún hay diferencias de cómo los diferentes actores lo conciben, por lo tanto, lo que se presenta en la Figura 3.3 es una propuesta del autor utilizando la metodología Canvas (Osterwalder y Pigneur, 2010). Luego se comentará sobre los principales elementos del diagrama y los que tienen relación con el presente proyecto.

Yentzen Media tiene dos principales segmentos de clientes: (1) los Lectores, quienes son los que acceden gratuitamente a los sitios web a consumir su contenido. Y segundo (2) son los Anunciantes que es toda aquella empresa o persona natural que pague por insertar su anuncio publicitario.

A cada segmento de cliente se le ha asignado un color (Figura 3.3), el cuál se utiliza dentro del diagrama para referenciar cada elemento al segmento correspondiente. Considerando lo anterior, se definen tres propuesta de valor:

1. **Actualidad y contenido:** Los lectores se enteran sobre lo que está ocurriendo en el mundo agrícola a través de las noticias proporcionadas. También se proporciona información que permite a los lectores orientar sus decisiones, como por ejemplo, reportes semanales del mercado con el avance de la temporada de las frutas frescas más relevantes del hemisferio sur.

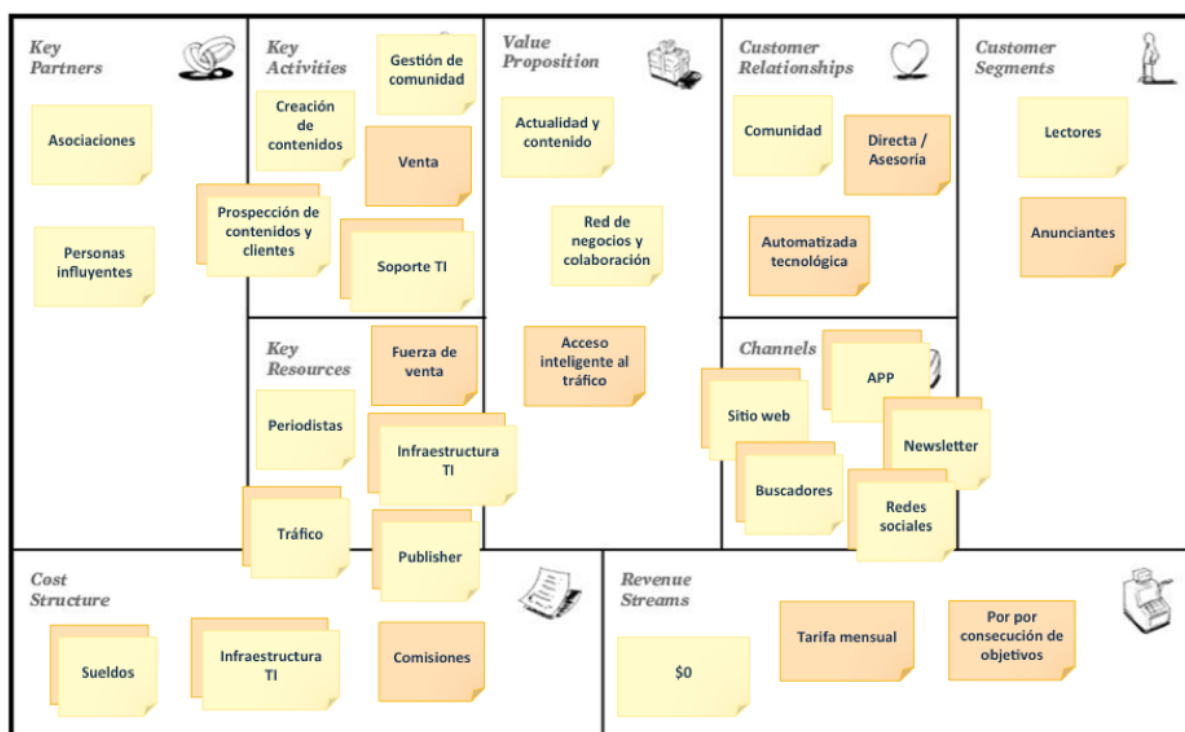


Figura 3.3: Modelo de Negocios propuesto para Yentzen Media

Una característica relevante, es que la mayoría de las noticias son de “factura propia” lo que quiere decir que son producto de la investigación periodística diaria del equipo. El resto es una combinación de contenido exclusivo que expertos y organizaciones publican en el sitio, y contenido proveniente de otros sitios web. Esto hace que el lector recurra en su visita.

En general, la información puede ser comercial, técnica y científica de interés. Los formatos que se maneja hasta el momento son: noticias, reportajes, estudios descargables, encuestas, data interactiva, videos, seminarios web.

2. **Red de negocios y colaboración:** Esta propuesta de valor no se encuentra activa hoy en día. Hace referencia a los mecanismos de interacción entre lectores con objetivos de negocio o aprendizaje. Los formatos de interés son: comunidades (sociales), biblioteca colaborativa, clasificados y directorio.
3. **Acceso al tráfico:** Esta propuesta de valor está orientada al segmento de Anunciantes, los cuales pueden insertar sus anuncios publicitarios en el sitio web y newsletter de los sitios web administrados por Yentzen Media. A la fecha, las posibilidades segmentación son a nivel de país y ubicación en la página web. Los formatos publicitarios que se operan son: banners y publrreportajes.

El presente trabajo mejora directamente esta propuesta de valor, al incorporar segmentación y asignaciones de anuncios utilizado semántica y el comportamiento del lector.

Finalmente, los mecanismos de ingreso son los siguientes:

1. Para los lectores es gratuito consumir los contenidos que se publican diariamente en el sitio web.
2. Para publicar un banner se establece una tarifa mensual, mientras que el servicio de publrreportaje y mailing, poseen una tarifa única.
3. Las tarifas flexibles por consecución de objetivos hoy no están activas, pero se espera que con el resultado de este trabajo, se esté un paso más cerca de implementarlas cuando la Dirección lo requiera.

Capítulo 4

Análisis Situación Actual

En este capítulo se detallarán los procesos de Yentzen Media utilizando la notación IDEF0 (Mayer, 1992), con las siguientes convenciones:

- Los procesos sombreados con negro indican que se profundizarán en ellos, puesto que el proyecto los afectará.
- Los procesos coloreados de amarillos indican el proceso está incompleto según el patrón propuesto por Barros (2009) o a juicio del autor, indicándose en cada caso.
- Las líneas rojas indican flujos inexistentes por según los patrones según el patrón propuesto por Barros (2009) o a juicio del autor, indicándose en cada caso.
- El patrón de arquitectura y estructura de procesos propuesta por Barros (2009) se ha modificado de la siguiente forma:
 - Se eliminó el proceso de “Mantenimiento de Estado” de los diagramas porque se asumirá que este opera “en segundo plano”, sin embargo, cabe señalar que este proceso tampoco está formalizado, centralizado, estandarizado ni disponible para el resto de los procesos.
 - Se eliminó el output de “Información de mercado” de los diagramas y los otros mensajes de la notación BPMN debido que se usará la notación formal IDEF0 (Mayer, 1992).

4.1. Arquitectura de Procesos

Aplicando los conceptos de macroprocesos desarrollados por Barros (2009), se instancia la arquitectura y estructura de los procesos de Yentzen Media. Al igual que las otras secciones, esta es una descripción sobre la “realidad observada”.

En la Figura 4.1 se presenta el nivel A-0 de la conceptualización de la compañía. Lo que finalmente es contenido e interacciones, es decir información en diferentes formatos que atraen a los lectores, quienes los consumen gratuita o no gratuitamente interactuando con ella y a la vez con otros actores. Para ello, como recursos habilitadores, se requieren sólo

de sistemas tecnológicos como WordPress, servicio de mailing, software de edición gráfica y una suite ofimática; lo que se relaciona directamente con las bajas barreras de entrada a esta actividad. Como ya se ha mencionado la cantidad y calidad de las interacciones entre usuario y entre usuarios y contenido no es la óptima, y esfuerzos se requerirán para avanzar hacia la visión propuesta.

Los inputs para generar contenidos son un poco más variados: información de mercado, recursos del mercado, requerimientos del cliente, necesidades del lector y material base. En una dinámica ideal, que aún no ocurre en la realidad, debería haber una coordinación entre los requerimientos del cliente y las necesidades del lector para generar los contenidos adecuados. El material base es el que se utiliza para elaborar o corregir el contenido pagado a insertar en los sitios web.

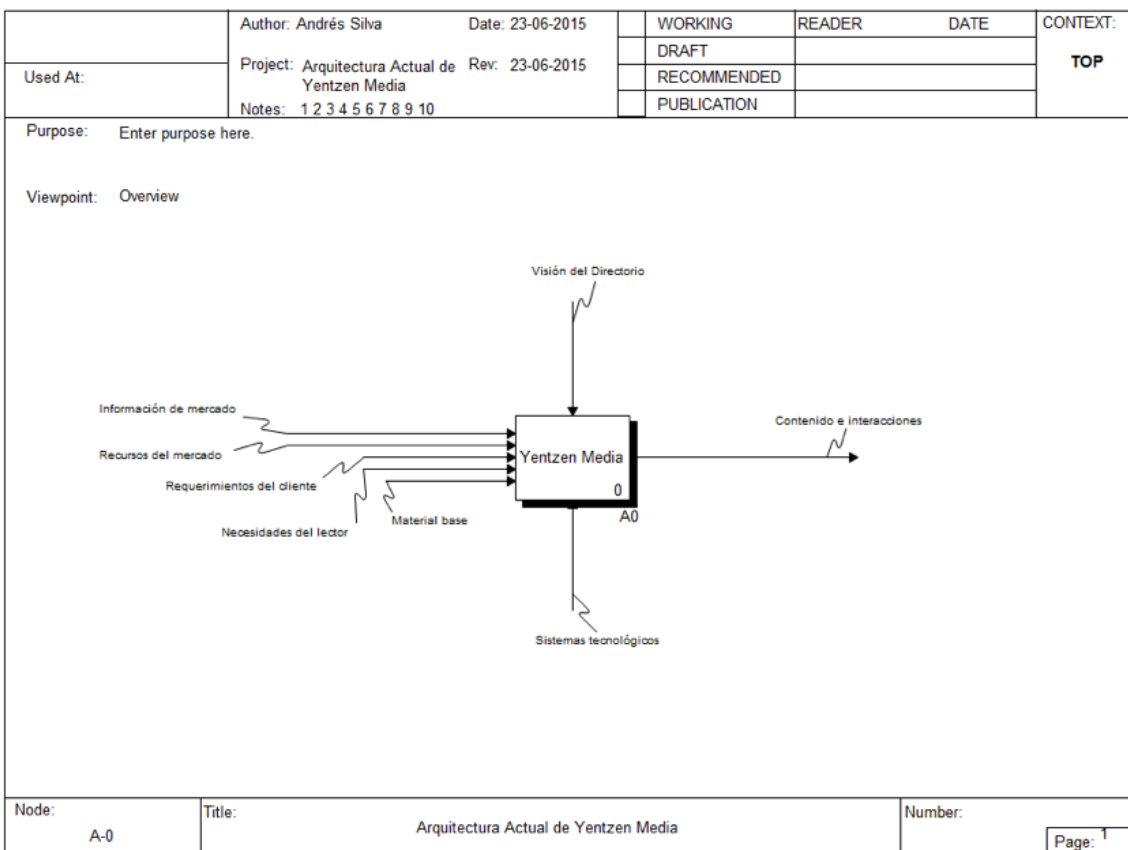


Figura 4.1: Nivel A-0 de Yentzen Media

En la Figura 4.2 se presentan los macroprocesos de Yentzen Media y las relaciones entre ellos. A continuación se describe cada uno de ellos, dónde la numeración indica el número de la macro en el diagrama IDEF0:

1. **Decisiones de Negocio:** Análogo al macroproceso 3 de “Planificación del negocio” propuesto por Brady (2009). Actualmente no hay una planificación formal del negocio, por lo que sólo se toman decisiones puntuales que no se pueden alinear con una visión u

objetivos estratégicos ya que estos no existen. La Dirección hoy en día está al tanto de este análisis y está tratando de mejorar la situación. Cabe destacar que efectivamente recibe información de mercado, ideas y resultados con ello se toman decisiones de una manera no estructurada ni formalizada.

2. **Desarrollo:** Análogo al macroproceso 2 de “Desarrollo de nuevas capacidades” propuesto por Barros (2009). Es un proceso relativamente nuevo en la organización que no está estructurado ni formalizado. Sin embargo, se están desarrollando nuevas capacidades de manera incipiente. Este proceso recibe también las ideas de las cadenas de valor. Enlazando con lo anterior, está claro que sin una planificación del negocio, no hay ni una sola garantía de que los proyectos ejecutados por esta área sean exitosos, ni que se inserten dentro de la dinámica correcta para maximizar los beneficios económicos y el valor de la empresa.
3. **Generación de contenidos:** Correspondiente al macroproceso 1 de “Cadena de valor” propuesto por Barros (2009). Esta primera cadena de valor consistente en la generación de información en diferentes formatos que permiten el entendimiento de la contingencia de la industria frutícola nacional e internacional. Hoy en día los contenidos se limitan únicamente a noticias y otros contenidos, como algunos reportes semanales de mercado. Éstos son proporcionados por proveedores, lo que está identificado como “Material base” en el diagrama. El análisis interior de esta macro revelará que tampoco hay análisis del comportamiento de los lectores o algún otro mecanismo que permita determinar cuáles son los contenidos relevantes a desarrollar.
4. **Servicio de publicidad digital:** Esta es la segunda cadena de valor que se coordina con la anterior, de forma que el output de esta cadena de valor (“contenido pagado”), son inputs de la anterior, por lo tanto, aquí hay un patrón de “cadenas de valor coordinadas” (Barros y Julio, 2011). El problema que hay con esta coordinación es que es meramente una instrucción, sin conexión alguna con el contenido ni los intereses con el lector. Entonces el resultado son un conjunto de anuncios molestos insertados en la página que no tienen necesariamente relación ni con el contenido ni con los intereses del lector. Adicionalmente, tampoco hay planificación ni generación de contenidos entorno a los anuncios activos, por lo que muchas veces, no se atrae el público objetivo de los anuncios para aumentar su conversión.
5. **Procesos de apoyo:** Este macroproceso tampoco queda ajeno a las “no definiciones” de la empresa, por lo que presentan ineficiencias y errores que producen que otras procesos no puedan funcionar adecuadamente.

4.2. Modelamiento detallado de procesos

4.2.1. Modelamiento IDEF0

En la Figura 4.3 se detalla el interior de la cadena de valor “Servicio de Publicidad Digital”, macroproceso que rediseña este proyecto. La primera observación, es que en la actualidad no

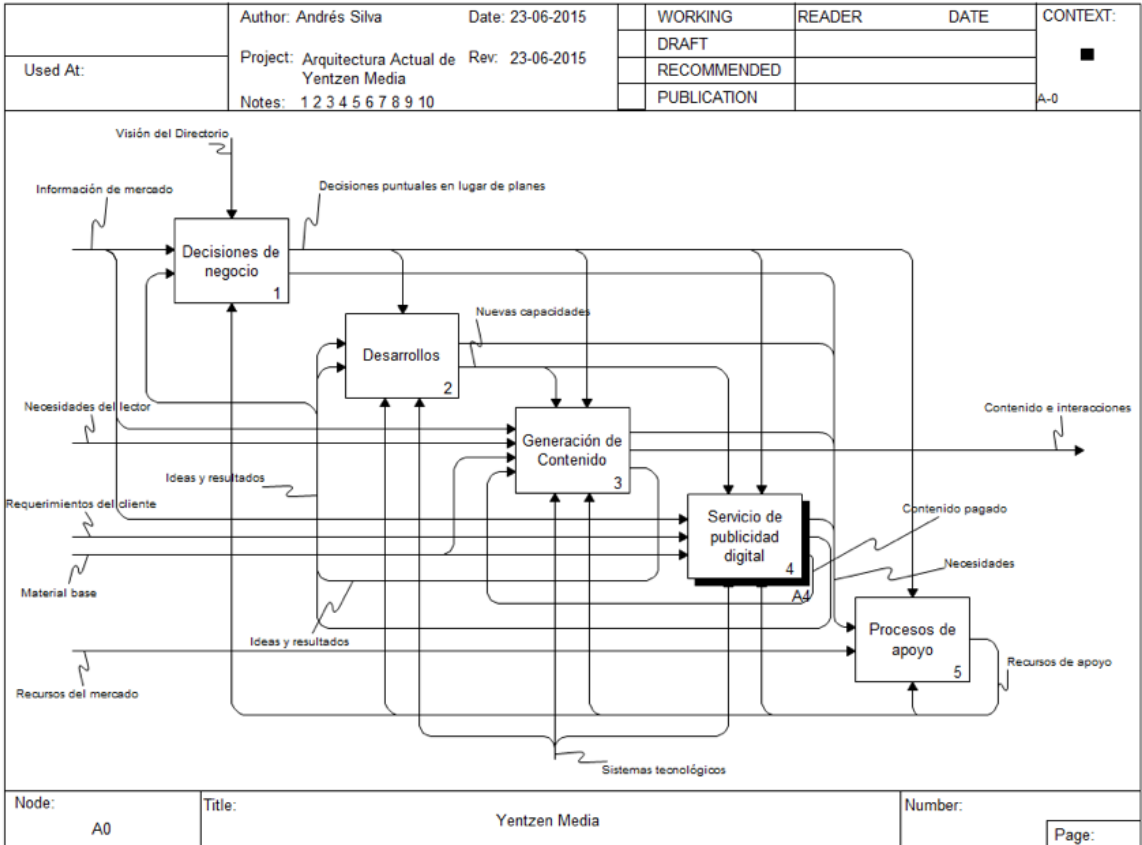


Figura 4.2: Macroprocesos de Yentzen Media

hay un proceso de gestión de proveedores puesto que no hay proveedores, sólo se recibe el material publicitario que el cliente envía y este es publicado. El no contar con este proceso ha sido una decisión no formal de la Dirección, ya que antes se contaba con un Diseñador que trabajaba adaptado o creando las piezas publicitarias para los clientes, sin embargo la Empresa ya no cuenta con esta capacidad. Sin duda, está produciendo complicaciones operaciones debido a que clientes están requiriendo este servicio.

El proceso de “Administración de la relación con el clientes” existe pero de forma incompleta según la estructura de procesos propuesta por Barros (2009). El proyecto buscará completar parte de este proceso.

La “Gestión de la producción y publicación” del material publicitario incluye su revisión, gestión de corrección y gestión de programación de publicación y gestión resolución de conflictos entre ventas, reservas y cotizaciones. Al momento de inicio del presente trabajo, sólo se contaba con un una herramienta de apoyo tecnológica muy básica para realizar éstas gestiones, conocida internamente como “System Banner”, por lo que este proceso era muy intensivo en administración (horas hombre). La actualización del System Banner incluye una primera versión para apoyar la gestión de éstas tareas, lo que se asemeja a un software BPMS hecho a medida de la empresa.

Finalmente, el proceso de “Producción, publicación y despliegue” se relaciona directamente el System Banner. Actualmente sólo soporta lógicas de segmentación del tráfico basados en la localización a nivel de país y posición en la página. Como se mencionó en los objetivos, el proyecto también trabajará en esta línea, proporcionando un sistema de recomendación que permitirá priorizar los anuncios a mostrar. Esto también podría ser utilizando en el mismo proceso pero de la cadena de valor de “Generación de contenidos”, pero para recomendar noticias u otra información. Ambas formas, se mejorarían las métricas del sitio y las campañas de los clientes.

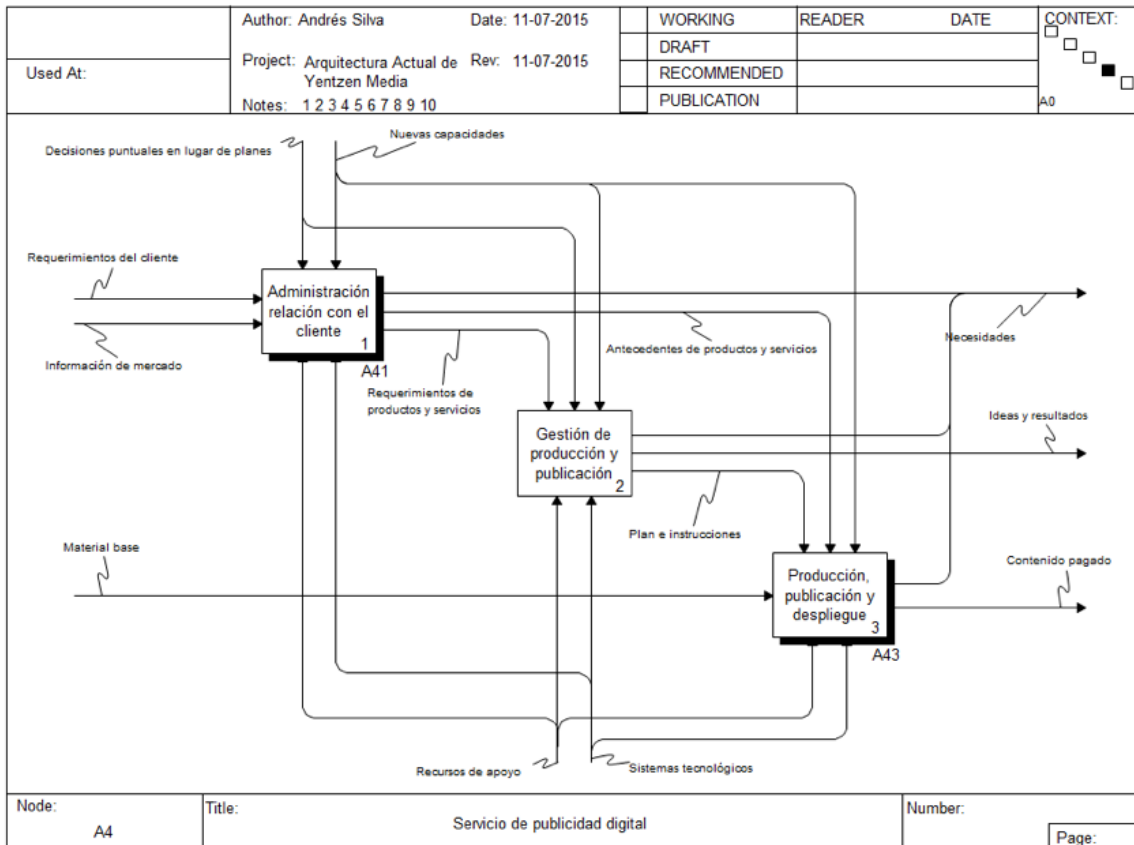


Figura 4.3: Detalle de la cadena de valor “Servicio de publicidad digital”

Profundizando lo anterior, en la Figura 4.4 se presenta el detalle del proceso “Administración de la relación con el cliente”. El proceso “Marketing y análisis de mercado” se encuentra coloreado de amarillo para indicar una alerta por incompletitud, mientras que las flechas rojas son flujos que hoy no existen y se consideran necesarios por el autor.

La inexistencia del flujo de “Instrucciones” desde “Marketing y análisis de mercado” a “Venta y atención al cliente” radica en que no existe una planificación de venta ni del negocio, sólo se realiza la acción de venta a discreción de los vendedores.

El flujo “Feedback” desde “Decidir satisfacción de requerimientos” a “Marketing y análisis de mercado”, se ha agregado pese a que no esté definida en los patrones propuestos por Barros

(2009), puesto que se considera importante que si no satisfacen los requerimientos, esto debe informarse para analizar nuevas oportunidades. El hecho de carecer un proceso y flujos para hacerse cargo de las demandas de los clientes está produciendo una brecha entre lo que los clientes quieren o esperan, con lo que el negocio obtiene como resultado. Esta es una razones de algunas de las razones de la no renovación de contratos, eventos que se estima que pueden aumentar en el mediano plazo si no se toman acciones correctivas.

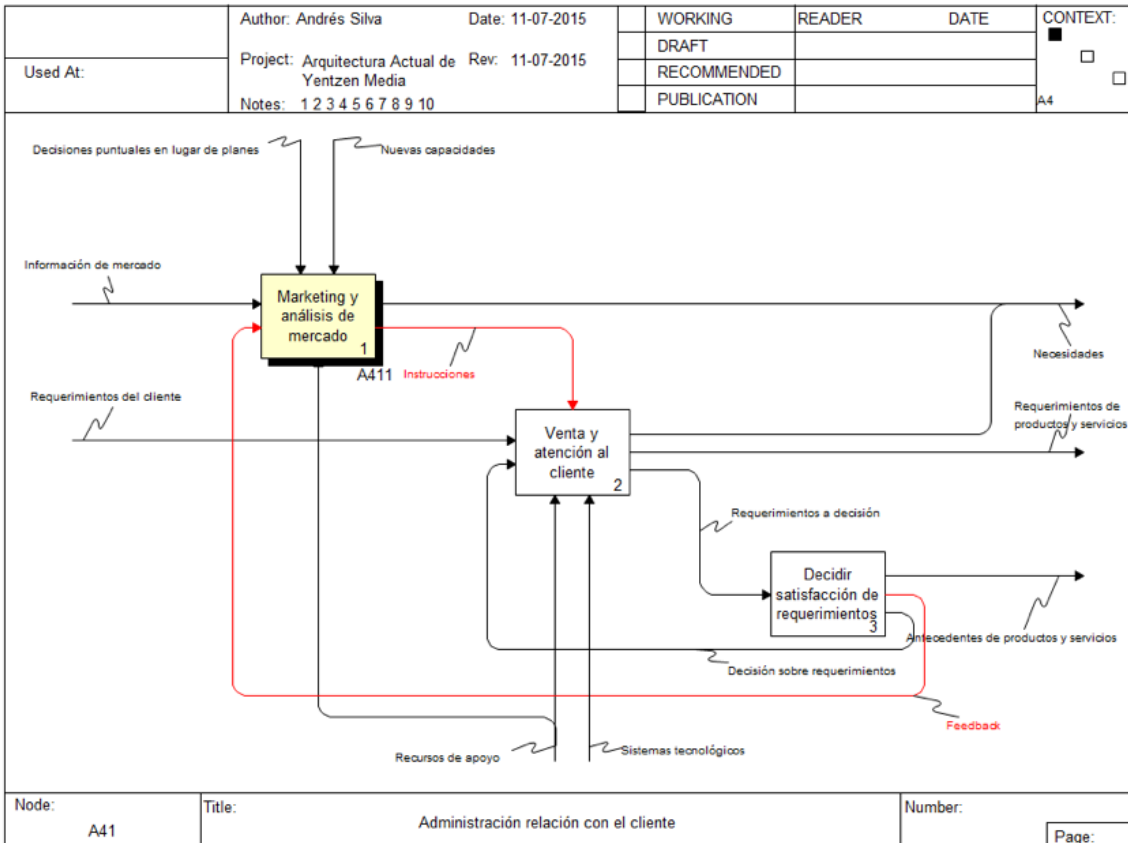


Figura 4.4: Detalle del proceso de “Administración de la relación con el cliente” de la cadena de valor “Servicio de publicidad digital”

Al profundizar en el proceso de “Marketing y análisis de mercado” (Figura 4.5), se evidencia las razones de la incompletitud señalada anteriormente, las cuales se comentan a continuación.

El proceso de “Introducción de nuevos productos” está incompleto y no es formal, lo que se explica porque no se tienen los mecanismos de feedback explicados anteriormente para determinar si se requieren ajustes en los productos. Únicamente no se consideró este proceso como inexistente es por la reciente creación de la aplicación móvil, la cuál está sin soporte en la actualidad.

El proceso de “Analizar el comportamiento de clientes y lectores” es inexistente en la actualidad, por lo tanto, todos los flujos que le relacionan tampoco existen. Esto se considera fundamental, ya que es la inexistencia de este proceso el que produce el efecto de “navegar

a ciegas”, por lo que las acciones de marketing se definen sin datos duros, sin modelos, sin inputs, sólo están siendo controladas por las decisiones puntuales de negocio como se comentó anteriormente. Esto afecta negativamente el proceso de “Planificación de ventas”, el que en todo el rigor debiera llamarse “Acciones de venta”, debido a que sólo existe un seguimiento comerciales de las acciones de venta que los vendedores toman según su criterio, sin la orientación una estrategia de la Dirección Comercial. Sin la comprensión del comportamiento de los lectores y clientes (y su cambio en el tiempo), ni esfuerzos en esta línea, no hay forma de mejorar la conversión de la lectura y las campañas publicitarias.

El proyecto aportará la base para para comenzar a entender éstos comportamientos en futuros proyectos internos.

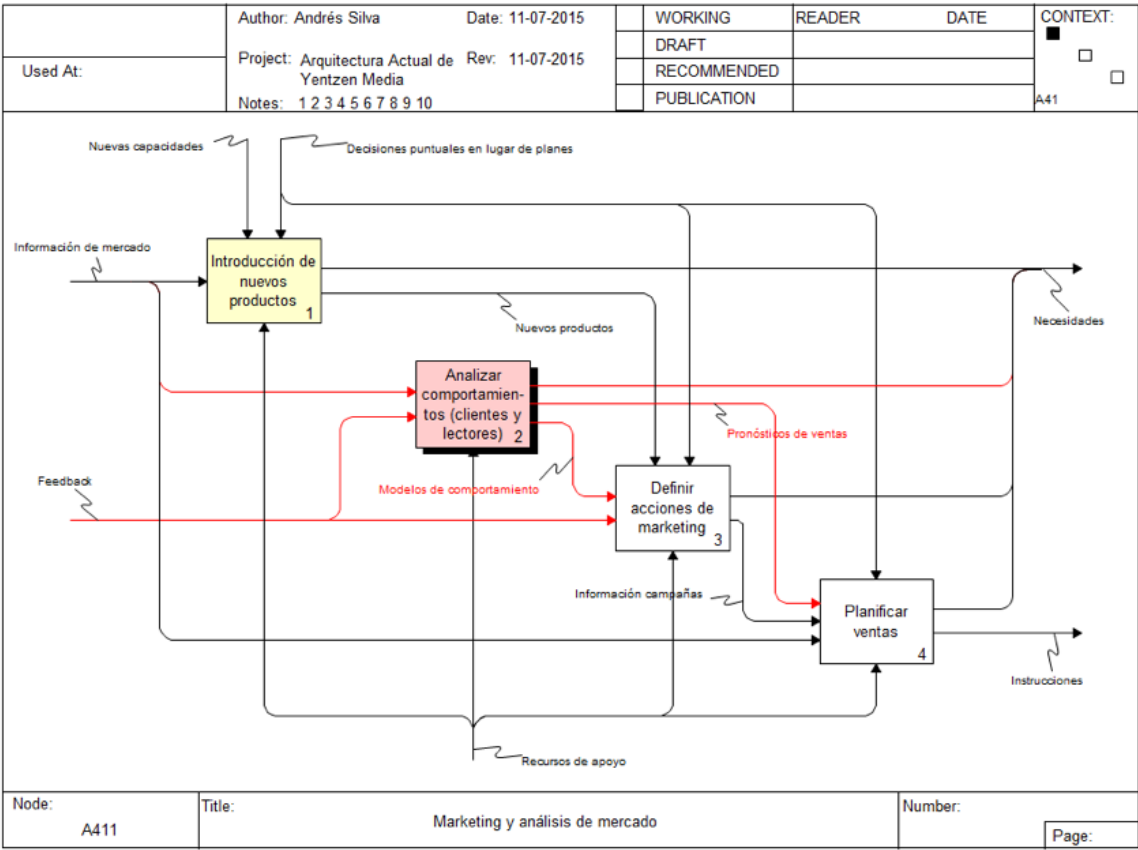


Figura 4.5: Detalle del proceso de “Marketing y análisis de mercado” de la cadena de valor “Servicio de publicidad digital”

Finalmente, en la Figura 4.6 se detalla el proceso de “Producción, publicación y despliegue” de la cadena de valor “Servicio de publicidad digital” (Figura 4.3). Este proceso existe actualmente, donde la principal tarea es la bajada y subida de banners al sitio web, tarea que se realiza manualmente en la plantilla del sitio web. El “System Banner”, que técnicamente debería ser un Ad Server, debería realizar este proceso automáticamente. La próxima versión de la herramienta pretende incorporar esta función básica que otros Ad Servers del

mercado tiene. Es por ello que se considera que no hay un sistema tecnológico que habilite adecuadamente este proceso.

Cabe destacar, aún cuando la nueva versión incorpore la subida/bajada automática de banners, aún queda la falencia de las lógicas que permitan hacer despliegue de los anuncios con segmentación “más inteligentes”, y no sólo a nivel de país como se realiza hoy.

Esta característica es lo que busca aportar el presente proyecto, característica que ya han solicitado algunos clientes y cuya demanda, según el autor, aumentará en el futuro cercano.

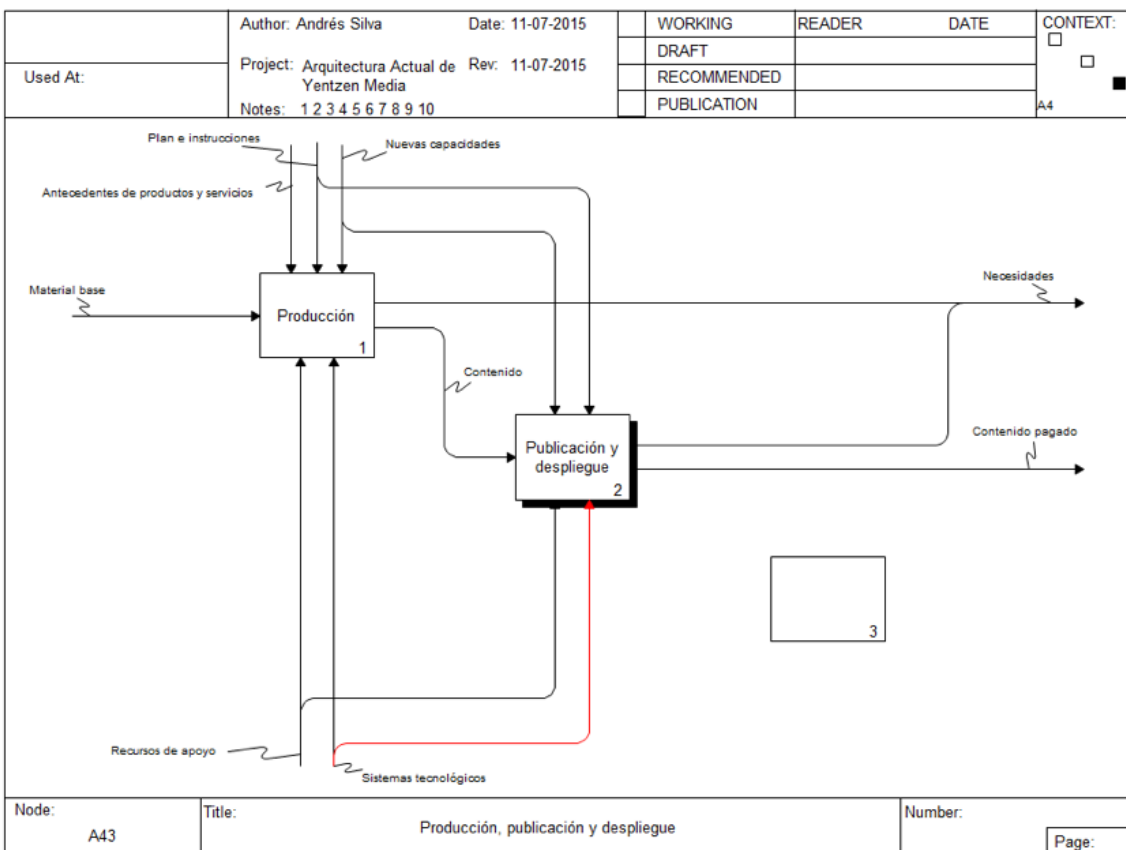


Figura 4.6: Detalle del proceso de “Producción, publicación y despliegue” de la cadena de valor “Servicio de publicidad digital”.

4.3. Diagnóstico de la situación actual

La actual arquitectura y estructura de procesos de la Empresa no permite hacerle frente a la problemática de negocio descrita en la sección 1.3. Parar ello, no sólo se debe trabajar en múltiples procesos e interacciones que hoy son incompletas o inexistentes, sino que también, se requiere que la Dirección revise su entendimiento del negocio y su gestión para liderar la

Empresa hacia una forma de trabajo más formal, ordenada y orientada hacia una estrategia definida que debe controlarse y cuestionarse periódicamente dada la naturaleza del negocio de los medios de comunicación digitales. De no realizar esto, la competitividad de la empresa se verá mermada en los próximos años, cuyos signos ya se comienzan a evidenciar con la no renovación de contratos y rechazo incrementado de propuestas.

Este proyecto no puede resolver todas éstas problemáticas descritas, dado su contexto académico y el alcance que ello supone. Sin embargo, es viable abordar el entendimiento del comportamiento de los lectores para la priorización de anuncios para mejorar los indicadores que los clientes toman en cuenta al momento de evaluar la renovación de un contrato.

El resultado de este entendimiento no sólo se espera que tenga un impacto en la renovación de contratos, sino que también sea el punto de partida para constituir los procesos y flujos bases para el proceso de “Marketing y análisis de mercado” con el que se podrán orientar tanto acciones comerciales como de generación de contenidos.

Sin embargo, Yentzen Media está en una posición muy favorable dado que las inversión publicitarias en medios online se encuentran al alza. Sin embargo, hay dos elementos que perjudican a la Empresa:

1. Pese a que no haya datos para comprobarlo, la experiencia en el negocio indica que la industria frutícola chilena aún es poco profesional en lo que refiere a la publicidad, especialmente la digital. Por lo que muchos clientes prefieren invertir en medios tradicionales como las revistas específicas, puesto que aún es una industria que está dominada por otra generación. Sin embargo, esto se revertirá con el tiempo.
2. La tendencia mundial de la publicidad online está basada en resultados (venta por impresiones, por clic, interacción u otro) donde el modelo de cobro implica mostrar resultados al cliente y después cobrar por ellos.

Sobre el primer punto, no se tiene más remedio que seguir trabajando en la educación de sus clientes, y con el tiempo, el mercado se ampliará más para la Empresa. Sin embargo, en el segundo punto, la empresa si tiene que comenzar a prepararse, foco del proyecto propuesto.

4.4. Cuantificación del problema u oportunidad

El equipo de Ventas no maneja una estadística fidedigna con las razones de las no renovaciones de contratos, pero se estima que las no renovaciones por causa de una campaña no efectiva está por el orden del 40 %, lo que es válido por campañas de 3 meses hasta de 1 año. Esto supone una pérdida mensual entre US\$1,000.00 - 3,000.00 mensuales por cliente.

Sin embargo, mientras que los competidores no han avanzado hacia modelos de marketing digital de resultados que permitan optimizar el presupuesto del cliente, Yentzen Media puede tomar la delantera posicionarse como un medio efectivo para invertir.

Capítulo 5

Propuesta de Diseño de Procesos

5.1. Direcciones de cambio y alcance

De acuerdo a las posibilidades de dirección de cambio (Figura 5.1), se explican a continuación las que este proyecto supone para la Empresa:

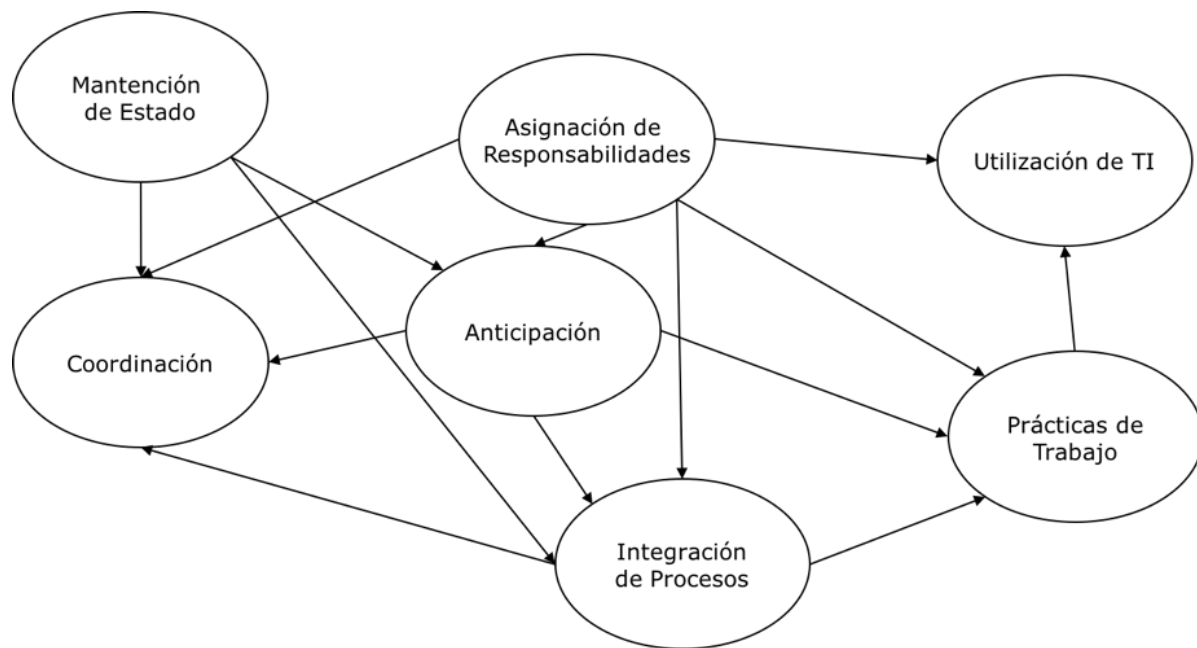


Figura 5.1: Posibilidades de Dirección de Cambio.

- **Coordinación:** Actualmente la modalidad de contrato de los anuncios consisten en la reserva del 100% del tráfico del sitio para una determinada posición y país, por lo que la posición nunca es compartida por más de un cliente. Esto simplifica la administración y coordinación de inicio/término de campañas para el equipo de Marketing.

Crear un nuevo espacio contextual, en el que los anuncios a desplegar se determinarán con cada instancia de visita, supone un nivel coordinación y administración muy

diferente, que manualmente no se podrá realizar.

El proyecto implica que Marketing mantendrá la administración de este nuevo espacio publicitario con la misma práctica de trabajo, pero la lógica que coordinará/asignará el anuncio a desplegar será compleja y automatizada, por lo tanto el cambio en este aspecto se dará en el “backend”.

- **Utilización de TI:** El proyecto implica un intensivo uso de tecnología y modelos de inteligencia de negocio, inclusive la utilización de base de datos no relacionales. Este sería el primer Sistema de Recomendación que la Empresa implementaría.
- **Integración de procesos:** La integración de procesos se realizará en la administración de las campañas en el “System Banner”. Para ello se requiere agregar sólo dos campos al modelo de datos de la línea de pedido¹ que Marketing deberá administrar. Con esto, se podrán habilitar/deshabilitar los anuncios automática para el sistema de recomendación.
- **Asignación de responsabilidades:** La mantención de los nuevos dos campos para las líneas de pedido es una nueva responsabilidad para Marketing, así como también la mantención del Sistema de Recomendación para el equipo de Desarrollo.
- **Mantención de Estado:** El proyecto implica la creación y mantención de perfiles de lectores, información que hoy no existen. Para ello, se integran los datos de varios sistemas diariamente.

5.2. Arquitectura de procesos To be

El autor propone que la Empresa debe avanzar hacia la arquitectura de procesos que se presenta en la Figura 5.2. El proyecto particularmente implicará implementar el 50 % del servicio compartido de “Análisis del comportamiento de lectores y clientes”, que sólo generará analizará el comportamiento de los lectores a través de la creación y mantención de perfiles.

Pero el presente trabajo, también sentará las bases para que la Empresa migre hacia una coordinación entre sus dos cadenas de valor. Esta sinergia permitirá aprovechar óptimamente las oportunidades de desarrollo de contenidos con la búsqueda de anunciantes de modo de proporcionar contenidos más relevantes para los lectores como también campañas publicitarias más ajustadas a los contenidos e intereses de los diferentes segmentos. Esto producirá un efecto positivo en la métrica de conversión. Esto se refleja en la arquitectura las instrucciones desde la cadena de valor “Servicio de publicidad digital” a “Generación de contenido”. Con esta información también será un input para el proceso de “Planificación de ventas” que se muestra en la Figura 5.3.

¹Una “Línea de pedido” corresponde a un anuncio a desplegarse en un tiempo y lugar determinado, por lo que una campaña publicitaria podría contar con varias “Líneas de pedido”, por ejemplo: un banner para el sitio web y un banner para el newsletter.

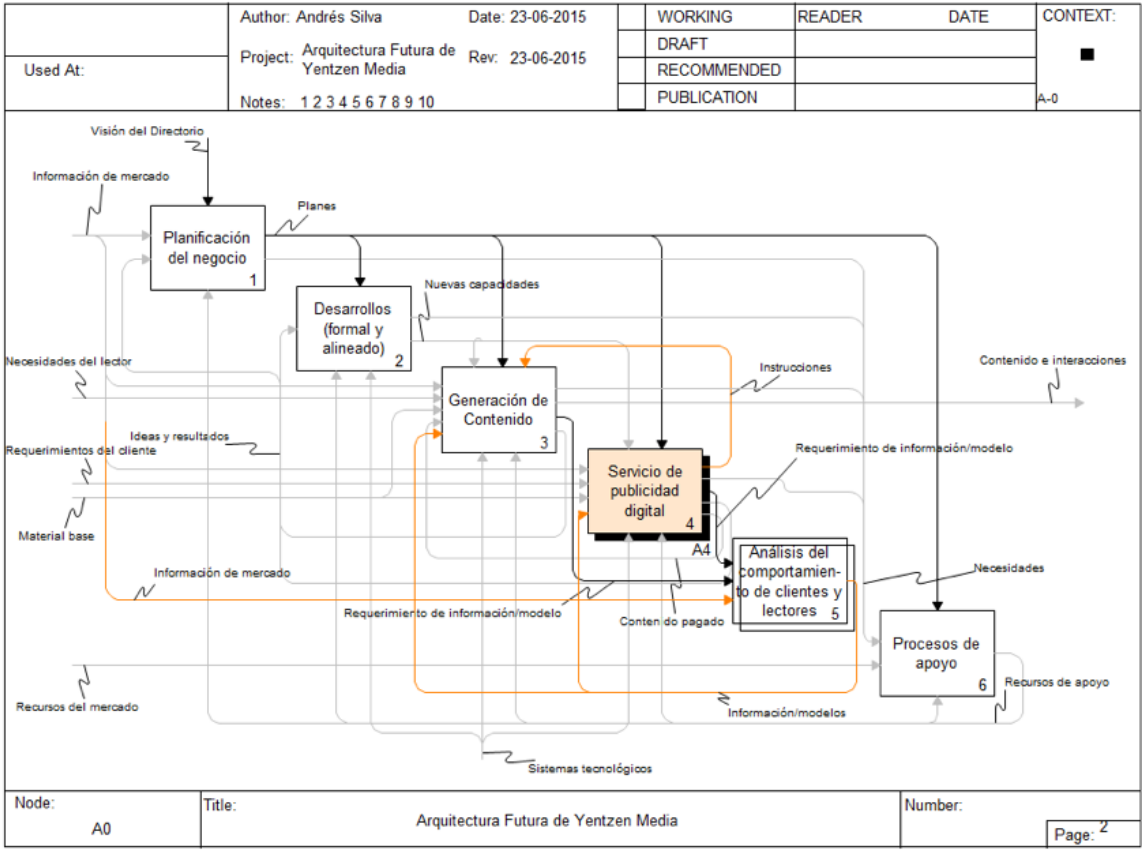


Figura 5.2: Arquitectura propuesta para Yentzen Media. El color naranja se destacan los flujos y procesos que serán agregados o modificados.

5.3. Diseño de procesos detallados To be

5.3.1. Modelamiento IDEF0

El proceso de “Publicación y despliegue” de los anuncios será el principal influenciado por el proyecto, que contempla la incorporación del recurso habilitador “Sistema de Recomendación” que permitirá gestionar el despliegue contextual de los anuncios, lógica que se incorporará en el “System Banner”.

5.3.2. Modelamiento BPMN

El proceso de entrega de anuncios se describe más adelante en la Figura 5.6 (sección 5.4 “Lógica del sistema de recomendación semántico”).

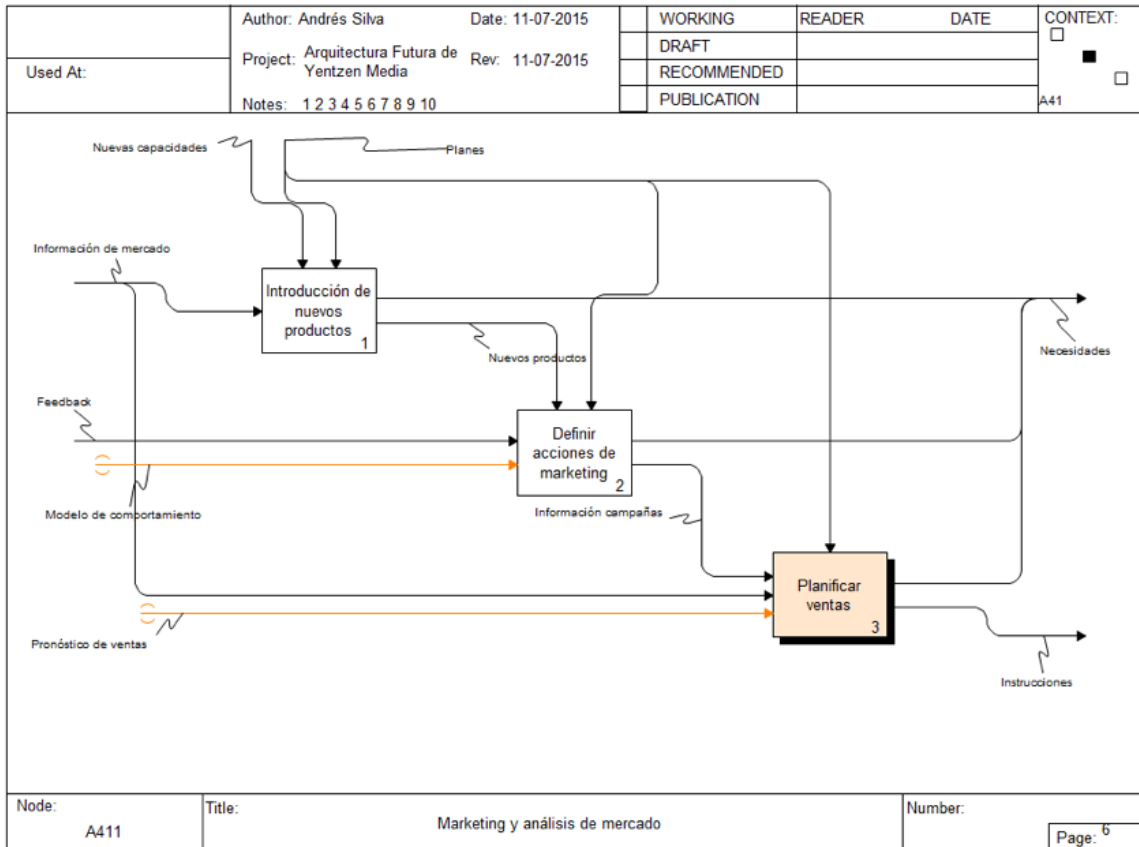


Figura 5.3: Estructura interna propuesta para el proceso de “Marketing y análisis de mercado” de Yentzen Media. Los nuevos flujos se encuentran destacados con color naranja, los cuales provienen del servicio compartido “Análisis del comportamiento del cliente y lector” (Figura 5.2).

5.4. Diseño de lógica de negocios

La lógica de negocios en este trabajo se compone de:

1. Diseño experimental
2. Lógica del sistema de recomendación semántico
3. Lógica del cálculo del vector del anuncio
4. Método de cálculo de perfiles

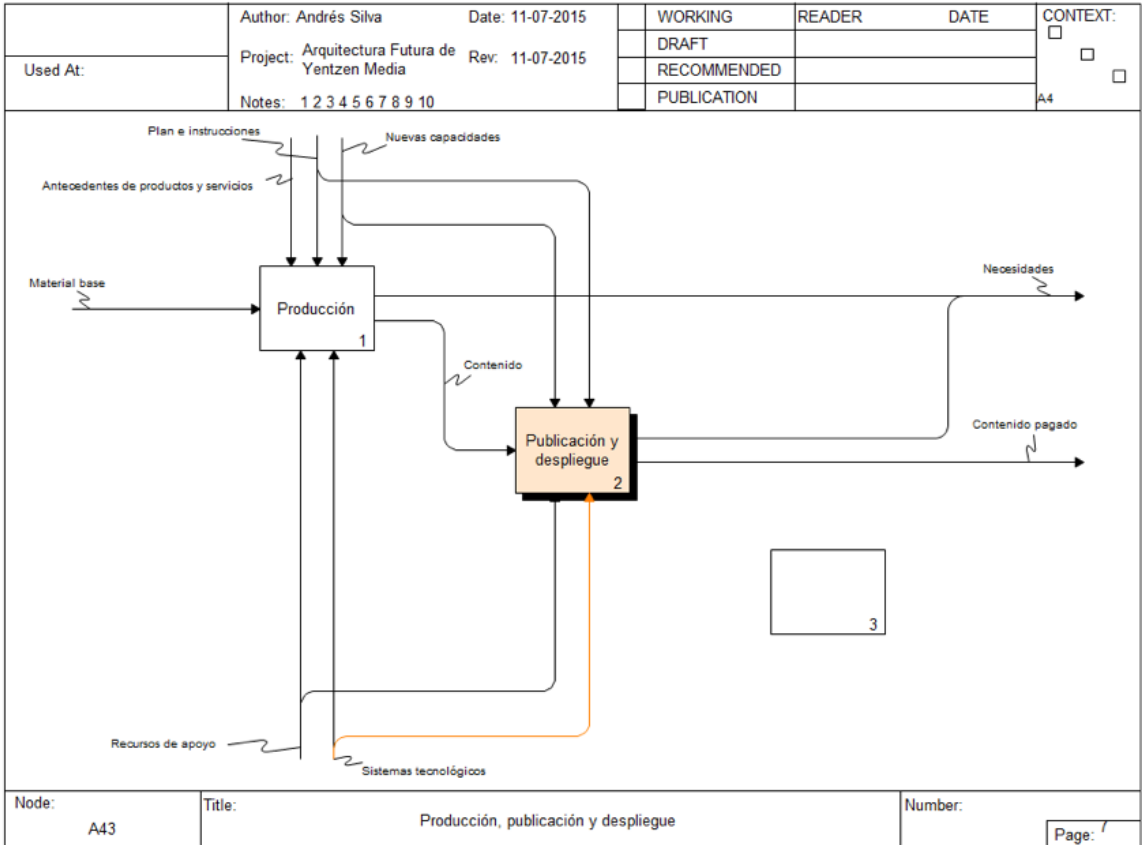


Figura 5.4: Estructura interna propuesta para el proceso de “Producción, publicación y despliegue” de la cadena de valor “Servicio de publicidad digital” de Yentzen Media. En naranja se destaca lo que será influenciado por el proyecto.

Diseño experimental

Identificación y enunciado del problema

Como se mencionó en la sección 1.3 “Problema u oportunidad identificada” el problema que representa este proyecto es del tipo de *exploración o caracterización de factores* que influyen en la interacción del usuario (lector) con los anuncios publicitarios dentro del ecosistema digital de Portalfruticola.com. Para ello, no sólo se debe determinar cuál es la forma más efectiva de presentar los anuncios, sino que también cuál es el mejor método para determinar la asignación adecuada. Por lo tanto, se realizará un experimento para determinar estadísticamente las mejores combinaciones, como también responder las interrogantes de investigación enunciadas en la sección 1.4.2 “Objetivos específicos”.

Objetivo del experimento

Determinar los factores, y su combinación, que mejoran la tasa de conversión de las campañas publicitarias en Portalfruticola.com en un nuevo espacio publicitario².

Variable de respuesta

La variable de respuesta a utilizar es el CTR (Fórmula 1.1, página 2), cuya caracterización se presenta en la Tabla 5.1.

Tabla 5.1: Caracterización de las variables de respuesta.

	Valor
Variable de respuesta	CTR
Unidad de medida	Porcentaje
Valor actual	0,24 %
Objetivo	Aumentar
Forma de medición	Fórmula 1.1

¹ Promedio mensual por campaña durante el 2015.

Selección de factores, niveles y rangos

En conjunto con el Equipo de Marketing se realizó una revisión de los factores que se creen que podrían influir en la variable objetivo. La Figura 5.5 se presenta un diagrama de causa-efecto con la clasificación de factores identificados.

De la Figura 5.5, y de acuerdo a la metodología de diseño experimental propuesta por Montgomery (2012), los *factores controlables* corresponden a los factores que interesa encontrar los niveles y combinaciones adecuadas. Los *factores a mantener constantes* son aquellos que no interesa determinar, en esta etapa, la influencia que tienen sobre la variable de respuesta y que además se pueden controlarse su valor, por ello se mantendrán a un valor constante para evitar aislarlos de los factores de diseño.

Los *factores ruido* son aquellos que se estima que perturbarán la variable respuesta en alguna medida, pero que son difícil de control, por lo que el ruido se aislará a través del análisis de bloques. Finalmente, los *factores incontrolables* son factores que afectarían la variable de respuesta, pero sobre los que no se tiene oportunidad de aislar, mantener constantes o controlar, dada su naturaleza o alcance de este proyecto.

En las siguientes secciones se tratarán éstos factores con mayor detalle.

²Las razones del nuevo espacio se explicaron en la sección 1.5 “Alcance”

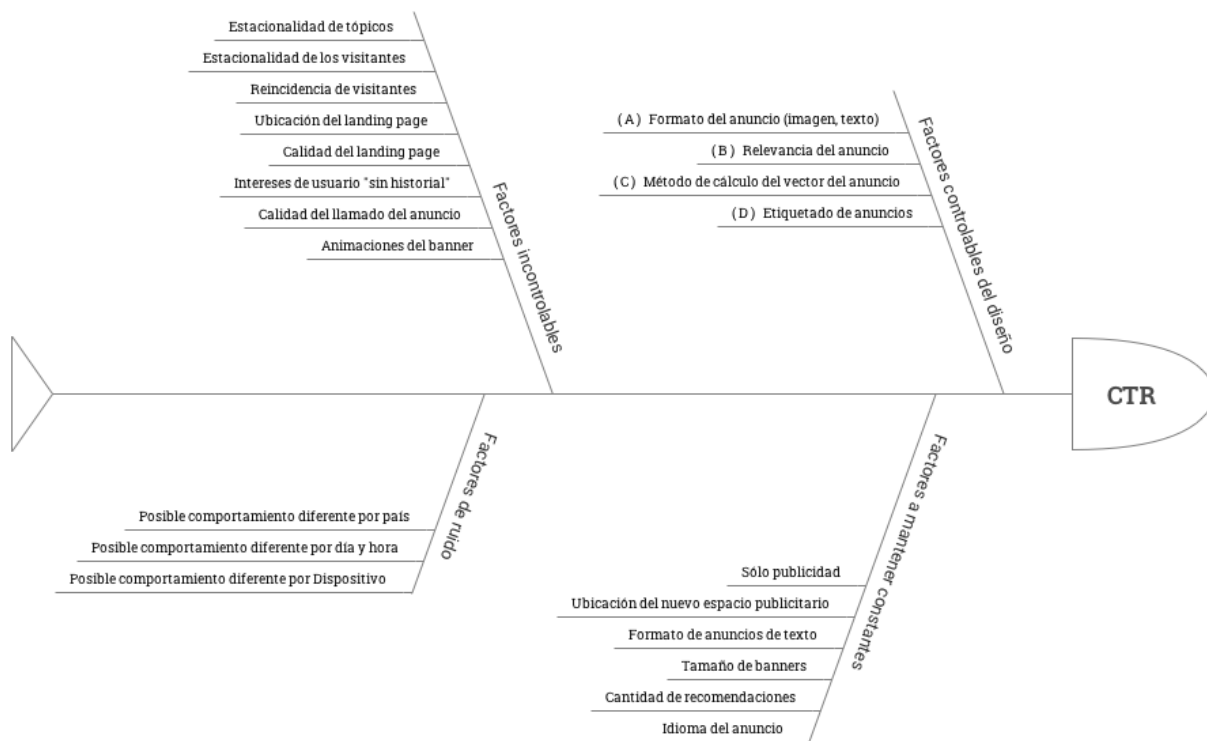


Figura 5.5: Diagrama de causa y efecto identificados que influyen en la variable de respuesta en el experimento.

Factores de diseño

Se han seleccionado 4 factores de control para el experimento, cuya descripción de niveles se presentan en la Tabla 5.2.

- A. **Formato:** es el tipo de anuncio, el cuál puede ser de texto o de imagen.
- B. **Relevancia semántica:** a través de una medida de similaridad semántica entre el anuncio publicitario con el par $\{\text{contenido}, \text{usuario}\}$ se computará un ranking de relevancia de anuncios. Se utilizarán los anuncios menos relevantes a la instancia para el caso “Mínimo”, y los más relevantes para el caso “Máximo”.
- C. **Etiquetado:** se refiere a si las categorías a las que pertenece el anuncio (ej: eventos, insumos, etc) se mostrarán junto con el anuncio para facilitar la interpretación del lector.

Los métodos de medición ha utilizar para determinar los niveles definidos anteriormente en la Tabla 5.2 se presentan en el Tabla 5.3.

Factores a mantener constante

Los factores a mantener constantes son:

Tabla 5.2: Niveles de los factores de diseño o variables de control para el experimento.

Factor de diseño	Nivel actual	Niveles para experimento	
		Bajo (-)	Alto (+)
Formato	Imagen	Texto	Imagen
Relevancia semántica	Desconocido	Mínimo	Máximo
Etiquetado	No aplica	Sin etiquetar	Etiquetado

Tabla 5.3: Método de medición de los factores de diseño o variables de control.

Factor de diseño	Método de medición
Formato	Visual
Relevancia semántica	Ranking en base a la similaridad del coseno
Etiquetado	Asignación Encargada de Marketing (Juicio Experto)

- **Sólo publicidad:** El sitio web no etiqueta los banners como anuncios publicitarios, pero si las noticias pagadas bajo la categoría “Publirreportajes”. Por ello, para seguir la línea del sitio web, en el nuevo espacio de recomendaciones, sólo se insertarán anuncios.
- **Ubicación del nuevo espacio publicitario:** Esta nueva zona se ubicará al final de las noticias, tanto en la versión Desktop como la versión Mobile.
- **Formato de anuncios de texto:** Los anuncios de texto serán de una sólo línea de no más de 60 caracteres de largo.
- **Tamaño de banners:** Sólo se utilizarán banners de 300x250 px.
- **Cantidad de recomendaciones:** Cuando las recomendaciones sean del tipo imagen se insertarán 2 y cuando sean de tipo twtxto se insertarán 6, tanto para la versión Desktop como para la versión Mobile.
- **Idioma del anuncio:** Los anuncios serán sólo en español.

Preferencias de diseño experimental

Dado que el experimento posee 3 variables, y cada una con dos niveles disponibles, método a utilizar es el correspondiente a experimentos factoriales 2^3 , con 8 corridas (Tabla 5.4). Adicionalmente, dado los factores de ruidos, se realizarán los siguientes bloques:

1. País: se presume un potencial comportamiento diferente por país, ya que deberían tener diferentes intereses dada las diferentes zonas productivas que cada uno posee.
2. Semana: dado el patrón de comportamiento por día y hora de la semana, se analizará la variable objetivo por semana para estandarizar las observaciones.
3. Dispositivo: dado que la versión mobile y desktop del sitio son diferentes, el objeto web que contendrá las recomendaciones tendrá una adaptación diferente por tipo de dispositivo, por lo tanto, los datos deben analizarse por separado.

Tabla 5.4: Corridas del experimento

Número de la corrida	Factor			Etiqueta de la corrida
	A	B	C	
1	-	-	-	(1)
2	+	-	-	a
3	-	+	-	b
4	+	+	-	ab
5	-	-	+	c
6	+	-	+	ac
7	-	+	+	bc
8	+	+	+	abc

Con respecto a las experimentaciones de prueba, se realizará una marcha blanca no para validar o calibrar los modelos, sino para probar que el sistema esté recolectando los datos de forma adecuada y a modo de prueba de estrés para la máquina virtual.

Técnicas de análisis y presentación

Para el análisis de los resultados y presentación de datos, se utilizarán las fórmulas definidas por Montgomery (2012) para los experimentos factores 2^3 .

Lógica del sistema de recomendación semántico

El algoritmo que se decidió utilizar para construir el sistema de recomendación de anuncios es el Latent Dirichlet Allocation (LDA). Con este algoritmo se construirá un modelo de semántico para describir cada noticia mediante un vector. La cantidad de elementos que este vector contiene depende la cantidad de tópicos que se elijan para al momento de entrenar el modelo, por lo tanto, es un modelo de predictivo semi-supervisado y una clasificación difusa. Es una clasificación difusa puesto que el número de cada posición del vector representa el grado de pertenencia de la noticia a cada tópico. Adicionalmente, este algoritmo permite predecir el vector para noticias que “no haya visto”.

Los vectores de las noticias se combinarán (ver sección 5.4 “Método de cálculo de perfiles”) para calcular perfiles individualizados por lector, y a su vez para determinar el vector de cada anuncio (ver sección 5.4 “Lógica del cálculo del vector del anuncio”).

La determinación del anuncio más relevante se realizará utilizando en base al vector de la instancia, la que se calculará utilizando la Fórmula 5.1. Cabe señalar que la decisión de asignar un 70 % de peso al perfil y un 30 % a la noticia actual, sólo obedeció a la intuición y el objetivo de hacer más relevantes los anuncios para los lectores de los que se pudo aprender su vector de intereses.

$$\text{Vector}_{\text{instancia}} = \begin{cases} \text{Vector}_{\text{noticia}}, & \text{if } \text{Vector}_{\text{noticia}} == \text{None} \\ \text{Vector}_{\text{perfil}} * 0,7 + \text{Vector}_{\text{noticia}} * 0,3, & \text{otherwise} \end{cases} \quad (5.1)$$

Luego, para computar el ranking de anuncios, el vector de instancia con el vector de cada anuncio disponible utilizando la similitud del coseno. De este modo se podrán seleccionar los anuncios relevantes y no relevantes dependiendo de la corrida del experimento. Este proceso se describe en la Figura 5.6.

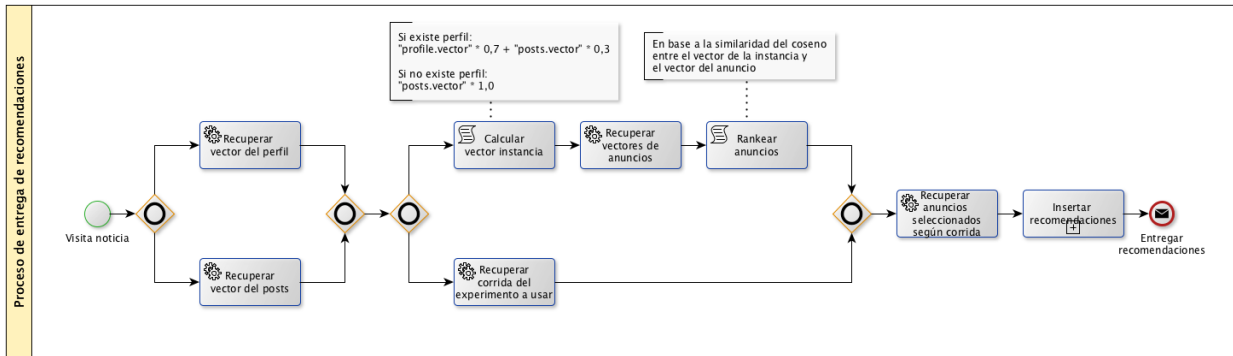


Figura 5.6: Proceso de recomendación de anuncios para el experimento.

Lógica del cálculo del vector del anuncio

El cálculo del vector del anuncio se hará en base al promedio de las 10 noticias más relevantes en las que teóricamente debería insertarse el anuncio. Para encontrar éstas 10 noticias se realizará una búsqueda de texto utilizando las keywords definidas para cada anuncio. Esta búsqueda es de tipo de término frecuencia (TF-IDF) en la que además se le asigna un peso específico dependiendo en los campos en que se encuentre la palabra clave en la noticia (ej: la noticia se considera más relevante si la palabra clave está en el título). De esta forma se obtiene un peso por noticia y permitiendo encontrar los mejores resultados.

Método de cálculo de perfiles

Para calcular los perfiles se realizará un análisis de los logs del sitio web. Para lograr identificar a la persona eficazmente, se consideran cuatro tipo de identificadores ordenados en orden de persistencia e importancia para el sitio:

1. **user_id**: corresponde al ID de la cuenta en Portalfruticola.com. Si el usuario navega iniciando sesión al sitio web, se obtendrá este ID para individualizar su tráfico.
2. **emai_id**: corresponde al ID del correo electrónico en la base de datos del newsletter. Si el usuario entra al sitio web por medio del newsletter, este ID será capturado.

3. **visitorId**: corresponde a un ID temporal alojado en un cookie en el navegador web. Este ID caduca después de 3 meses o puede ser eliminado por el usuario.
4. **IP**: los logs proporcionan la IP de conexión del usuario.

El ID utilizado para agregar el comportamiento web del usuario será en el orden expuesto en la lista anterior. Por lo tanto, también se debe considerar un proceso para agregar y consolidar perfiles. Por ejemplo cabe la posibilidad de que al usuario sólo se le puede asociar el visitorID, pero luego de unos días se registre el sitio por lo que estará disponible su userID. Entonces deben consolidarse los perfiles.

Dada las tecnología que se utilizará para capturar los datos y procesarlos, se realizará una actualización de los perfiles todos los días por la noche. Entonces, cada día el sistema debería poder entregar mejores recomendaciones debido a su aprendizaje.

Capítulo 6

Propuesta de apoyo tecnológico

6.1. Especificación de requerimientos

6.1.1. Requerimientos funcionales

1. El sistema de recomendación debe ser capaz de individualizar el tráfico a nivel de persona.
2. Debe soportar las diferentes corridas del experimento definido.
3. Debe guardar todos los datos de las recomendaciones realizadas y si tuvieron alguna interacción para su posterior análisis estadístico.

6.1.2. Requerimientos no funcionales

Se definen los siguientes requerimientos no funcionales:

Requerimientos Ergonómicos

1. Las recomendaciones deben poder clicarse fácilmente con el mouse o el dedo para el caso de dispositivos móviles.

Requerimientos de Disponibilidad

1. Tanto el experimento como la solución final implementada debe estar disponibilidad el 99,9% del tiempo.

2. La caja de recomendaciones debe estar disponible sólo para insertarse en los posts de tipo “noticias”.

Requerimientos de Interfaz

1. Para el usuario, las recomendaciones las verá directamente en el sitio web.
2. Para el administrador, no se considera una interfaz para la carga de datos, esto se hará de forma manual hasta que termine el experimento.

Requerimientos de Desempeño

1. La entrega de la recomendación debe entregarse en menos de 1 segundo al momento que un usuario acceda a una noticia.

Requerimientos de Entorno

1. Las recomendaciones deben ser proporcionadas en formato HTML para que puedan ser desplegadas en WordPress.
2. La aplicación se desarrollará en Google App Engine y la base de datos será MongoDB en una máquina virtual Linux.

Requerimientos de Entrenamiento

1. No es necesario entrenamiento alguno de parte del lector para su uso.
2. El administrador del sistema de recomendaciones si debe conocer en profundidad su funcionamiento para realizar las actualizaciones pertinentes y también la carga de datos de nuevos anuncios.

Restricciones de diseño

1. Este nuevo espacio publicitario se ubicará después de cada noticia.
2. Dado el espacio disponible para este nuevo espacio publicitario, en el caso de las imágenes, sólo se podrán poner dos y para los anuncios de tipo texto 6. Esto es válido para la versión mobile y desktop.

6.2. Arquitectura tecnológica

La Figura 6.1 muestra la arquitectura tecnológica diseñada para el funcionamiento de la aplicación.

6.3. Diseño de la aplicación

6.3.1. Casos de Uso

La Figura 6.2 ilustra los casos de usos y su relación por los diferentes actores.

6.3.2. Diagramas de Secuencia

Obtener recomendaciones

La Figura 6.3 muestra el diagrama de secuencia para el caso de uso “Obtener recomendaciones”.

Actualizar colección posts

La Figura 6.4 muestra el diagrama de secuencia para el caso de uso “Actualizar colección posts”, tarea que se ejecuta cada día durante la noche. Esta tarea es relevante puesto que predice los vectores de cada noticia para preparar las predicciones del día.

Actualizar perfiles

La Figura 6.5 muestra el diagrama de secuencia para el caso de uso “Actualizar perfiles”, tarea que se ejecuta cada día durante la noche.

Insertar nuevo anuncio

La Figura 6.6 muestra el diagrama de secuencia para el caso de uso “Insertar anuncio”, el cuál es un proceso manual que se realiza directamente en la base de datos a través de la consola. El trabajo posterior a la experimentación considera un proceso automático de inserción del anuncio desde el System Banner.

6.3.3. Diagramas de Clases

La Figura 6.7 muestra el diagrama de Clases del proyecto.

Diagrama de actores

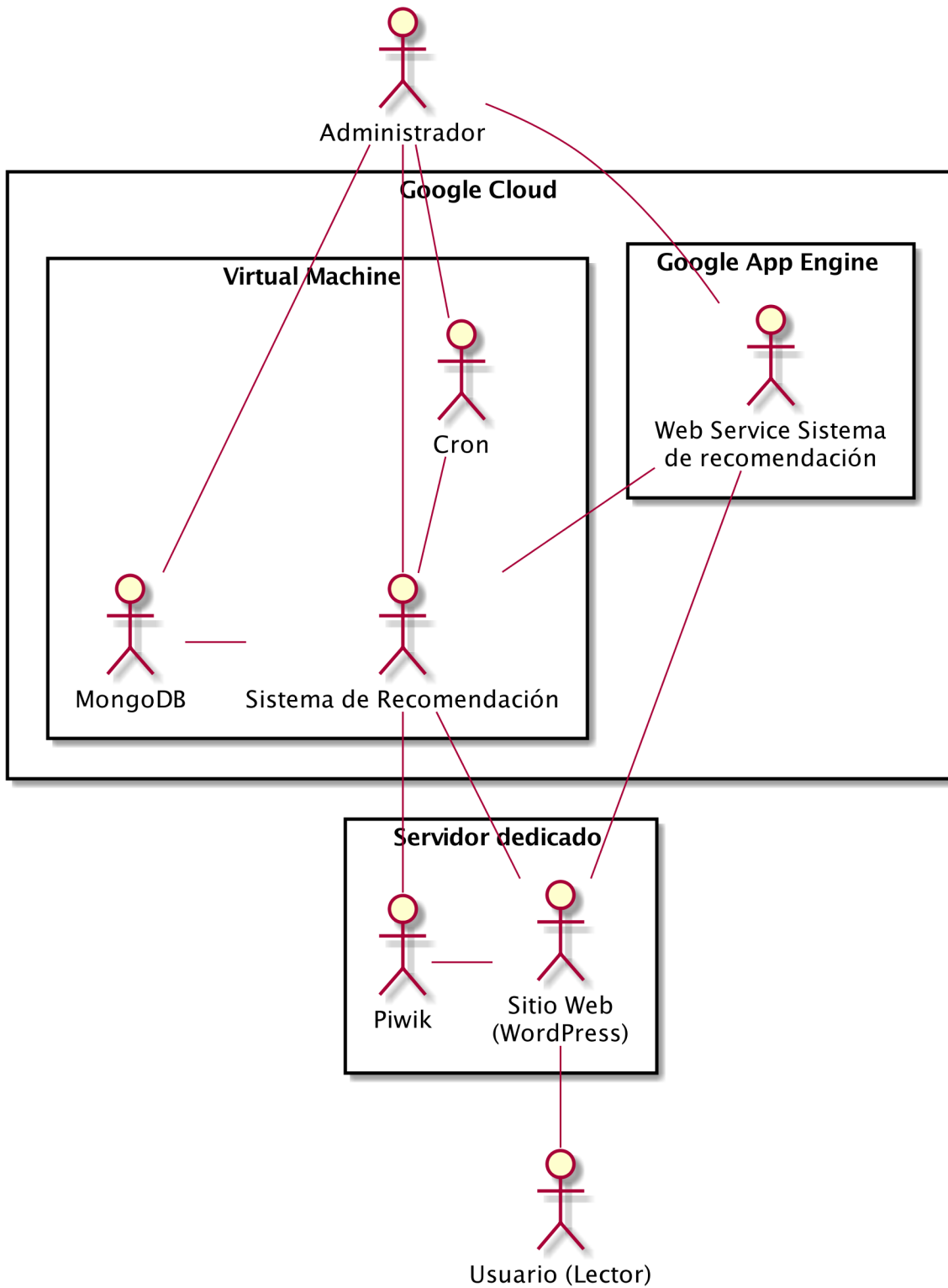


Figura 6.1: Diagrama de actores de la arquitectura tecnológica.

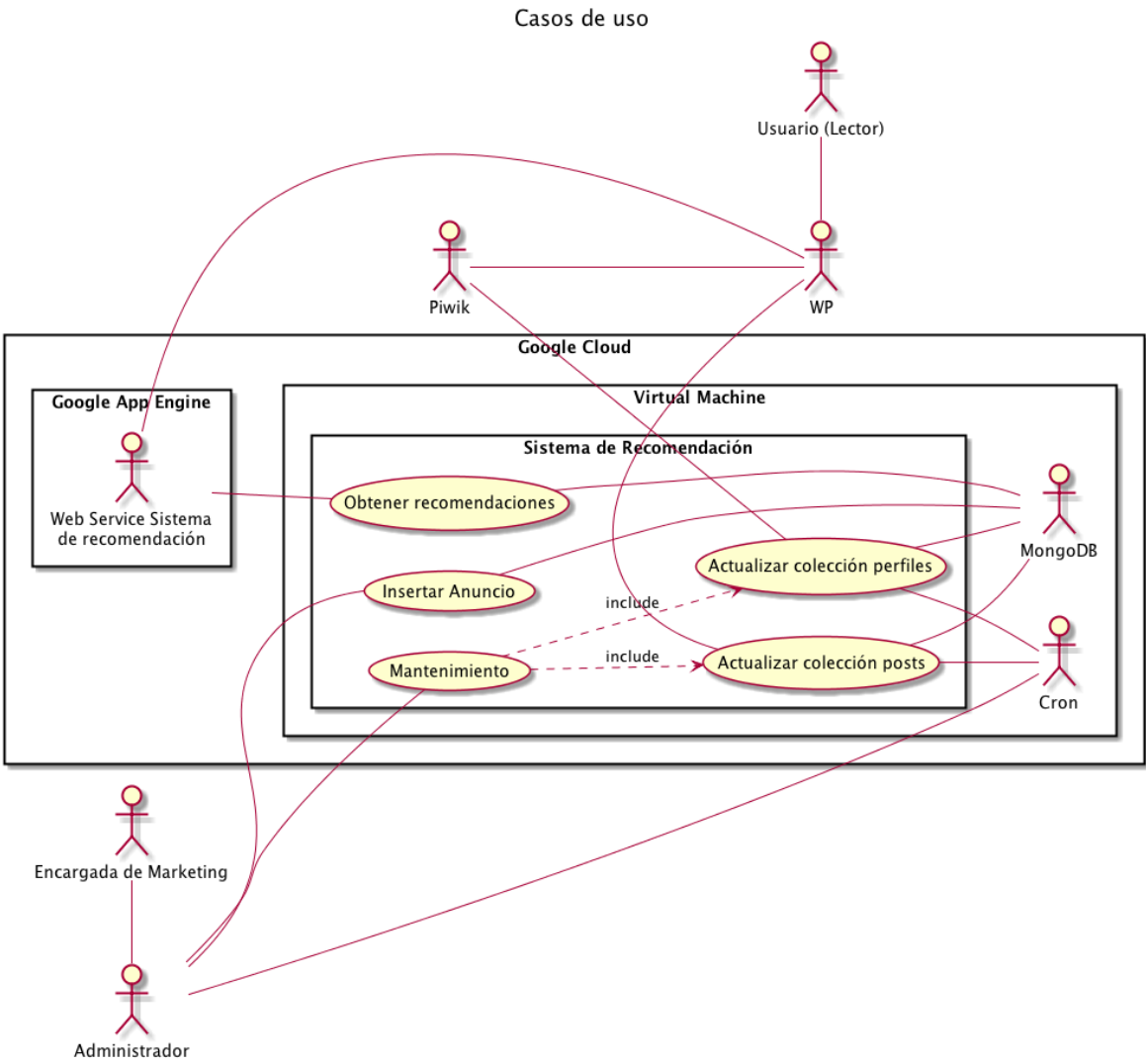


Figura 6.2: Diagrama de casos de uso.

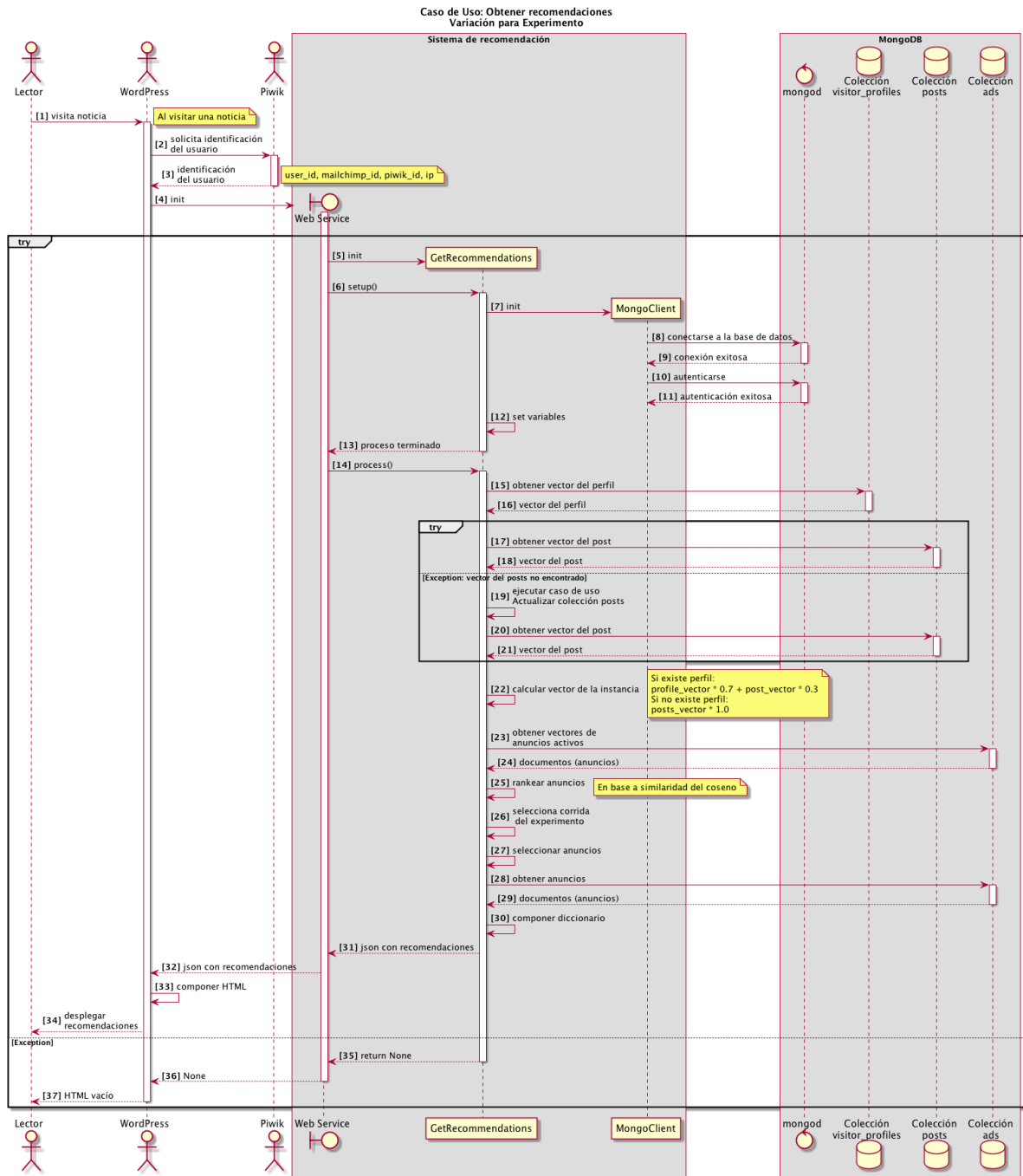


Figura 6.3: Diagrama de Secuencia del caso de uso “Obtener recomendaciones”

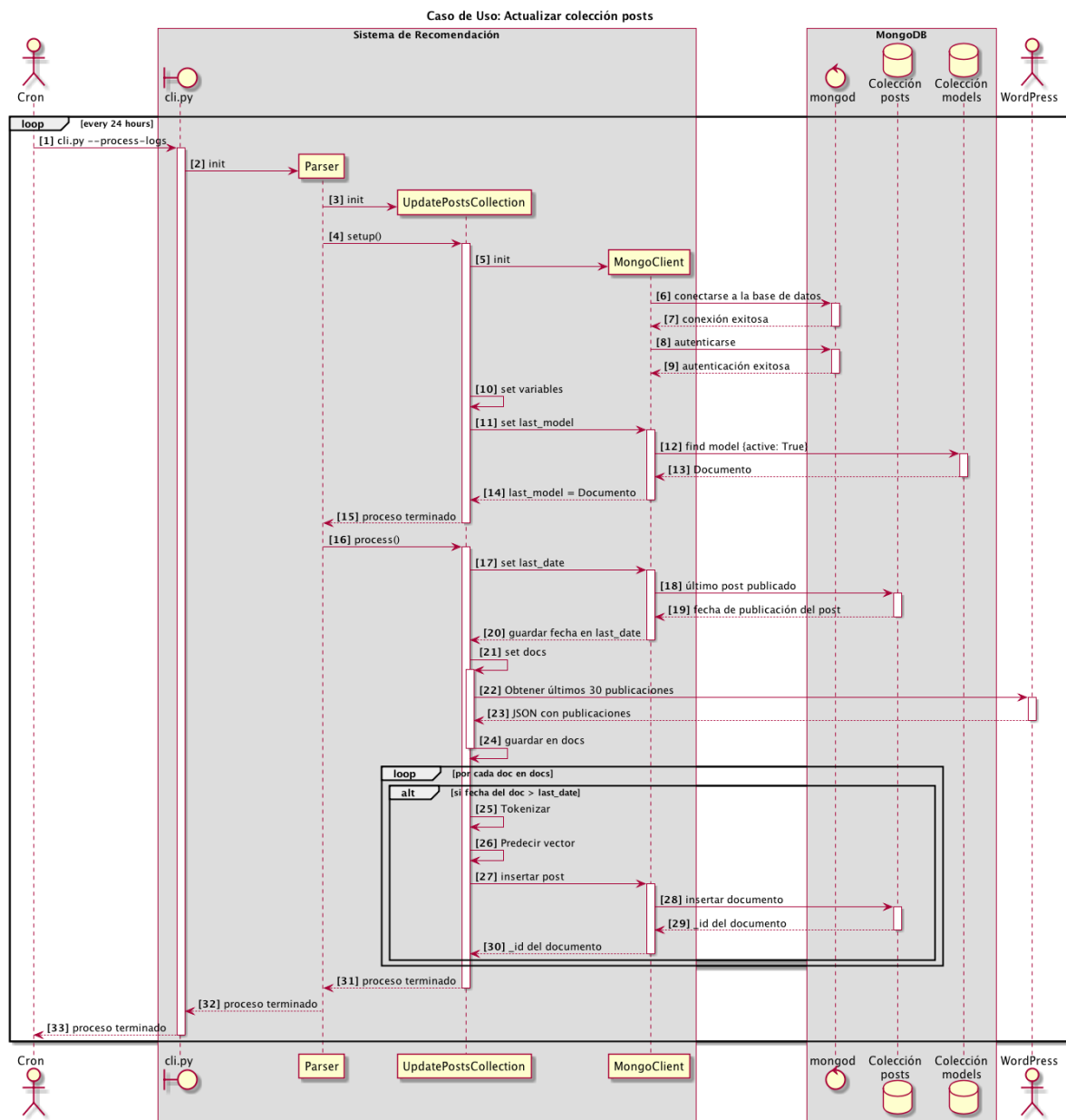


Figura 6.4: Diagrama de Secuencia del caso de uso “Actualizar colección posts”

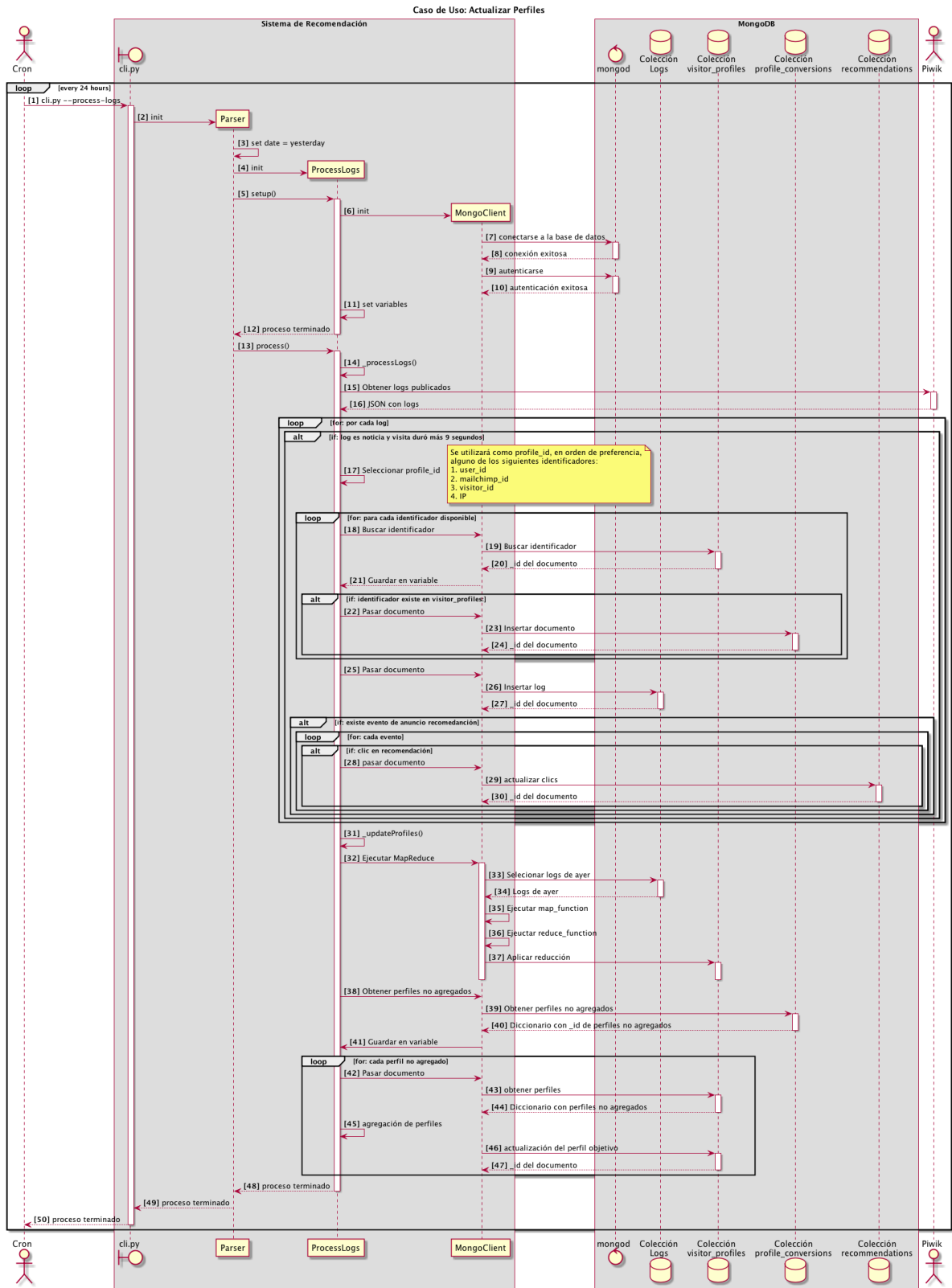


Figura 6.5: Diagrama de Secuencia del caso de uso “Actualizar perfiles”

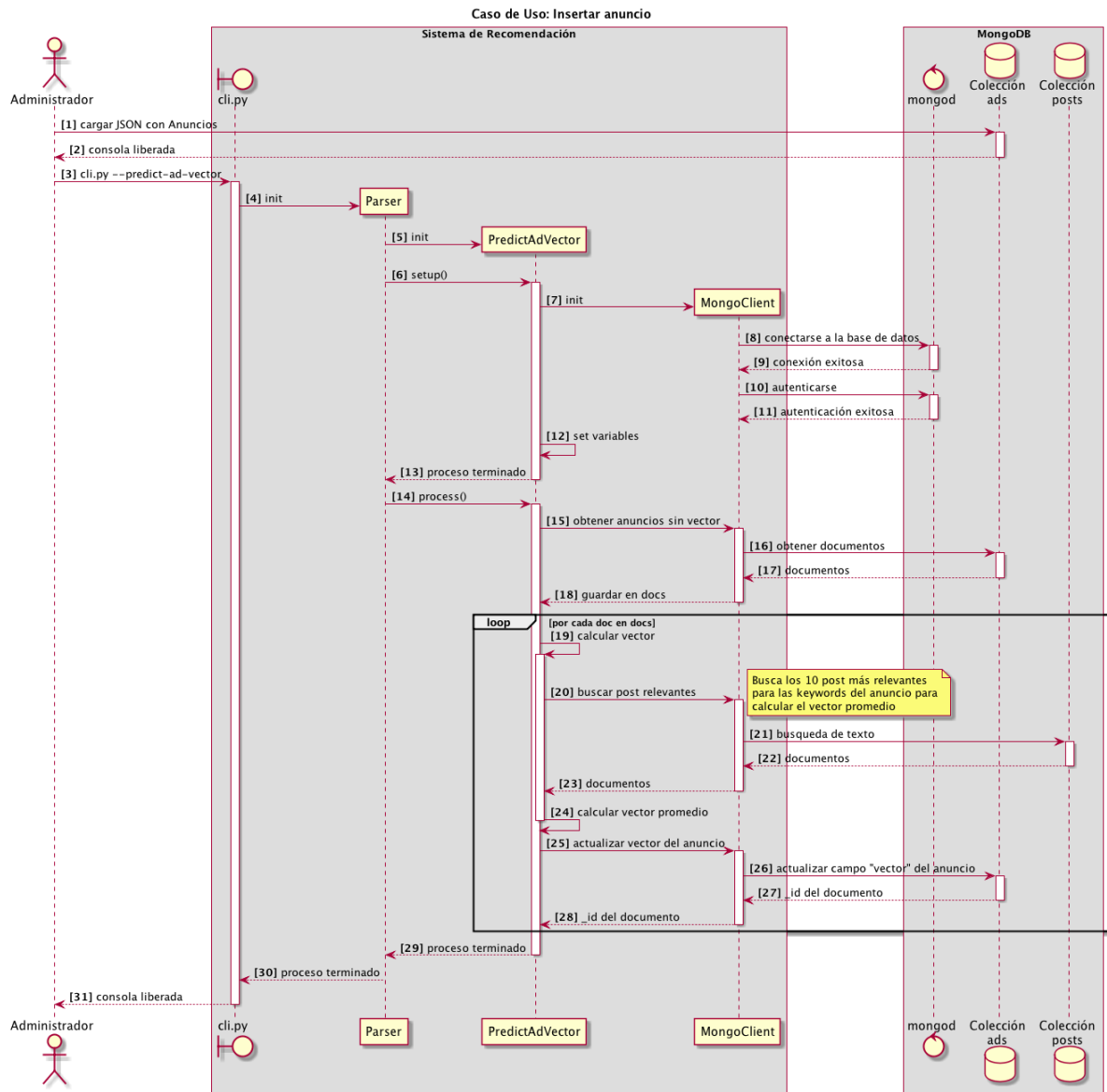


Figura 6.6: Diagrama de Secuencia del caso de uso “Insertar anuncios”

Diagrama de Clases

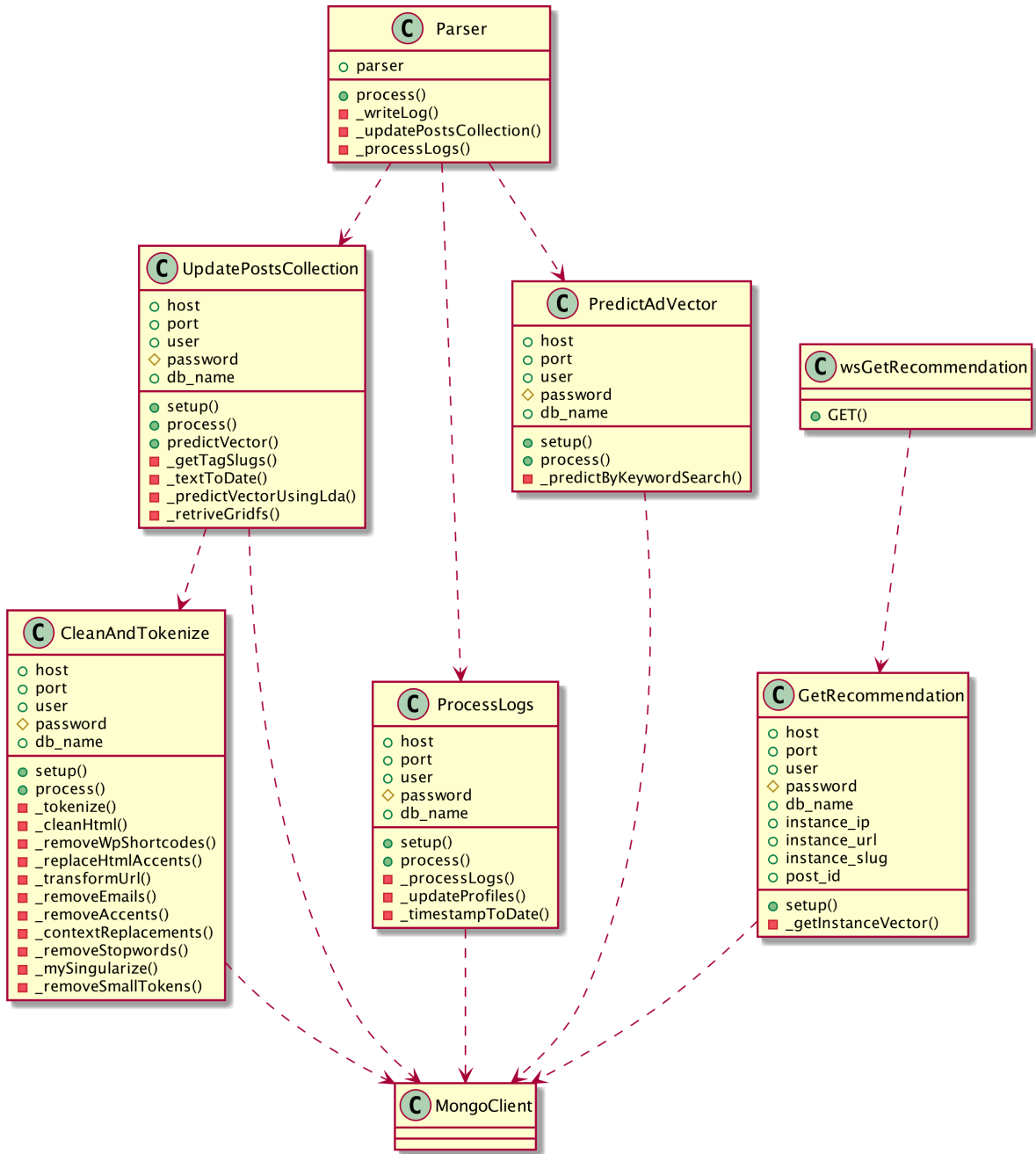


Figura 6.7: Diagrama de Clases

Capítulo 7

Evaluación del Proyecto

7.1. Definición de beneficios y costos

Tanto para la definición de los beneficios y costos de la evaluación del proyecto consideran los parámetros y supuestos presentados en la Tabla 7.1, los cuales se comentan a continuación:

- La evaluación se hará en pesos chilenos, para ello se consideró un valor dólar y de UF que para la fecha en que se hace la evaluación.
- La “Tasa de clientes nuevos por trimestre”, “Tasa de renovación promedio sin proyecto”, “Tiempo promedio de campaña”, “Costo promedio de auspicio mensual por cliente”, son supuestos basados en mediciones del negocio, dado que la empresa no proporcionó datos exactos en éstos puntos. La “Inversión promedio por cliente campaña”, se obtuvo utilizando los supuestos anteriores.
- La “Tasa promedio de crecimiento de Páginas Vistas (PV)” consideró la situación optimizada de crecimiento del 1 %, puesto que la tasa real calculada (Cuadro 8) es igual a -1,9 %.
- Como parte de los beneficios, se consideró el supuesto de que este nuevo espacio publicitario pudiera soportar la venta de campañas por CPM, por lo que sería un beneficio adicional del proyecto. Se estimó un precio de venta de venta CPM igual a \$9.475 considerando las impresiones generadas sólo para noticias durante 1 año a contar de Julio 2016 (Tabla 7.2). Para este valor, se consideró cobrar un 30 % adicional puesto la especificidad del público objetivo el sitio web atrae.

Adicionalmente, se estableció no se generarían ventas por este modelo el primer trimestre de operación (después del experimento), y que luego de este tiempo, sólo se lograrían vender el 20 % del total de las páginas vistas en este espacio con el pricing CPM.

- Las estimaciones se realizarán considerando como unidad de tiempo mínima un trimestre, puesto que el tiempo mínimo promedio de duración de las campañas publicitarias es de 3 meses. Adicionalmente, se considerará como horizonte de evaluación 2 años.

Considerando lo anterior, en la Tabla 7.3 se muestra la estimación de los beneficios y

costos incrementales atribuibles al proyecto.

Tabla 7.1: Parámetros y supuestos utilizados en la evaluación financiera del proyecto.

	Tipo	Valor	Unidad
Cambio dólar (24 de mayo de 2016)	Dato	\$694	CLP / USD
Valor UF (24 de mayo de 2016)	Dato	\$25.975	CLP
Tasa de clientes nuevos por trimestre	Supuesto	3	
Tasa de renovación promedio sin proyecto	Supuesto	40 %	
Aumento renovación por proyecto	Supuesto	5 %	
Tiempo promedio de campaña	Supuesto	3	Mes
Costo de auspicio mensual promedio	Supuesto	\$693.890	
Inversión promedio por cliente campaña	Supuesto	\$2.081.670	
Valor hora Data Scientist mantención modelo	Dato	\$25.975	
Depreciación software (lineal 2 años)	Supuesto	8	Trimestre
Impuesto a la renta anual	Dato	21 %	
Impuesto a la renta trimestral	Cálculo	5,3 %	
Tasa de crecimiento promedio de PV	Cálculo	1 %	
Promedio de PV mensual	Cálculo	95.202	
Porcentaje de venta por CPM	Supuesto	20 %	
Precio de venta CPM	Cálculo	\$9.475	

Tabla 7.2: Cálculo del precio de venta por CPM.

	Valor
Tasa de crecimiento mensual	-1,9 %
Promedio impresiones mes	95.202
Costo promedio campaña mes (US)	\$1.000
eCPM (US)	\$10,5
eCPM (CLP)	\$7.289
Sobrecargo nicho	30 %
CPM venta (dólares)	\$13,7
CPM venta (CLP)	\$9.475

Tabla 7.3: Estimación de ingresos y costos incrementales atribuibles al proyecto.

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
INGRESOS								
Ingresos por renovaciones								
Clientes iniciales	10							
Clientes nuevos	3	3	3	3	3	3	3	3
Total clientes potenciales	13	16	19	22	25	28	31	34
Abandonos (sin proyecto)	8	10	11	13	15	17	19	20
Abandonos (con proyecto)	7	9	10	12	14	15	17	19
Estimación de total clientes (sin proyecto)	5	6	8	9	10	11	12	14
Estimación de total clientes (con proyecto)	6	7	9	10	11	13	14	15
Ingresos potenciales	\$27.061.710	\$33.306.720	\$39.551.730	\$45.796.740	\$52.041.750	\$58.286.760	\$64.531.770	\$70.776.780
Pérdidas por abandonos (sin proyecto)	\$16.653.360	\$20.816.700	\$22.898.370	\$27.061.710	\$31.225.050	\$35.388.390	\$39.551.730	\$41.633.400
Pérdidas por abandonos (con proyecto)	\$14.571.690	\$18.735.030	\$20.816.700	\$24.980.040	\$29.143.380	\$31.225.050	\$35.388.390	\$39.551.730
Ingresos estimados (sin proyecto)	\$10.408.350	\$12.490.020	\$16.653.360	\$18.735.030	\$20.816.700	\$22.898.370	\$24.980.040	\$29.143.380
Ingresos estimados (con proyecto)	\$12.490.020	\$14.571.690	\$18.735.030	\$20.816.700	\$22.898.370	\$27.061.710	\$29.143.380	\$31.225.050
Diferencia ingresos con y sin proyecto	\$2.081.670	\$2.081.670	\$2.081.670	\$2.081.670	\$2.081.670	\$4.163.340	\$4.163.340	\$2.081.670
Ingresos por ventas CPM								
Impresiones	285.607	288.463	291.348	294.261	297.204	300.176	303.178	306.210
Ingresos adicionales por venta CPM	\$-	\$546.647	\$552.113	\$557.634	\$563.210	\$568.843	\$574.531	\$580.276
Ingresos incrementales								
Ingreso incrementa con proyecto	\$2.081.670	\$2.628.317	\$2.633.783	\$2.639.304	\$2.644.880	\$4.732.183	\$4.737.871	\$2.661.946
COSTO FIJO								
Mantención modelo (h)	16	16	16	16	16	16	16	16
Mantención modelo (\$)	\$415.608	\$415.608	\$415.608	\$415.608	\$415.608	\$415.608	\$415.608	\$415.608
Google App Engine y Storage	\$55.511	\$55.511	\$55.511	\$55.511	\$55.511	\$55.511	\$55.511	\$55.511
Total costo fijo	\$471.119	\$471.119	\$471.119	\$471.119	\$471.119	\$471.119	\$471.119	\$471.119
COSTO VARIABLE								
Comisiones Ventas CPM (10%)	\$-	\$54.665	\$55.211	\$55.763	\$56.321	\$56.884	\$57.453	\$58.028
Comisiones Renovaciones (5%)	\$104.084	\$104.084	\$104.084	\$104.084	\$104.084	\$208.167	\$208.167	\$104.084
Total costo variable	\$104.084	\$158.748	\$159.295	\$159.847	\$160.405	\$265.051	\$265.620	\$162.111

7.2. Flujo de caja

Para el cálculo del flujo de caja se consideran los siguientes supuestos y cálculos:

1. La inversión a realizar es el experimento.
2. El horizonte de evaluación es de 8 trimestres porque el tiempo máximo promedio de duración de un contrato es de 1 año y el mínimo promedio es de tres meses. De esta manera, se podrían tener en consideración al menos 1 renovación de contratos de 1 año.
3. El Capital de Trabajo requerido son sólo horas hombre de la supervisión de la Encargada de Marketing y del trabajo del Diseñador Gráfico para la actualización de los instrumentos de venta. El resto del trabajo no es incremental para el proyecto, porque se utilizarán las mismas instancias de ventas y reuniones para comunicar y vender el nuevo servicio. No se considera horas hombres adicionales en la administración del System Banner puesto que el sistema final extraerá la información automáticamente del mismo sistema. El campo de “palabras claves” será el único campo adicional que la Encargada de Marketing deberá procurar rellenar, cuya información es proporcionada por el cliente, por lo que se considera marginal.
4. La tasa de descuento a utilizar se calculó utilizando la metodología CAPM (Formu-la 7.1). Los valores utilizados para el cálculo de la tasa de descuento, y su fuente, se presentan en la Tabla 7.4. Para el tasa libre de riesgo, se utilizó el valor de los bonos en pesos a 2 años (horizonte de evaluación del proyecto). La tasa de retorno de mercado se extrajo de Bloomberg utilizando el índice IGPA, y el Beta se extrajo de la web Damodaran, para la industria “Advertising”, el que fue actualizado en Enero 2016.

$$\begin{aligned} \text{CAPM} &= K_{LR} + (K_m - K_{LR})\beta \\ &= 3,87 + (9,62 - 3,87) * 1,08 \\ &= 10,08 \end{aligned} \tag{7.1}$$

Dónde:

K_{LR} : Tasa libre de riesgo
 K_m : Rentabilidad de mercado
 β : Beta del activo

Tabla 7.4: Valores de las variables para calcular la tasa de descuento a través del método CAPM.

Símbolo	Valor	Fuente
K_{LR}	3,87	Banco de Chile
K_m	9,62	Bloomberg
β	1,08	Damodaran

Considerando lo anterior se calculó el flujo de caja privado (Tabla 7.5) y los indicadores económicos. Se obtuvo un VAN de \$11.528.966 y TIR 40 %, por lo que el proyecto debería ejecutarse.

Tabla 7.5: Flujo de Caja Privado. Todos los valores son en pesos chilenos.

	Quarter								
	0	1	2	3	4	5	6	7	8
Ingresos por Venta	2.081.670	2.628.317	2.633.783	2.639.304	2.644.880	4.732.183	4.737.871	2.661.946	
Costo Variable Total	-104.084	-158.748	-159.295	-159.847	-160.405	-265.051	-265.620	-162.111	
Costo Fijo Total	-471.119	-471.119	-471.119	-471.119	-471.119	-471.119	-471.119	-471.119	
Gastos generales									
Depreciación activo fijo tangible									
Amortización activo fijo intangible	-860.625	-860.625	-860.625	-860.625	-860.625	-860.625	-860.625	-860.625	-860.625
Intereses									
Perdida por venta activo fijo									
Perdida por venta capital de trabajo									
Utilidad Antes de Impuestos	645.844	1.137.827	1.142.747	1.147.718	1.152.737	3.135.394	3.140.514	1.168.099	
Impuesto	33.907	59.736	59.994	60.255	60.519	164.608	164.877	61.325	
Utilidad Después de Impuestos	611.937	1.078.091	1.082.753	1.087.462	1.092.219	2.970.785	2.975.637	1.106.774	
Depreciación activo fijo tangible									
Amortización activo fijo intangible	860.625	860.625	860.625	860.625	860.625	860.625	860.625	860.625	860.625
Intereses									
Perdida por venta activo fijo									
Perdida por venta capital de trabajo									
Flujo de Caja Operacional	1.472.562	1.938.716	1.943.378	1.948.087	1.952.844	3.831.410	3.836.262	1.967.399	
Inversión Activo Fijo	-6.885.000								
Inversión capital de Trabajo	-200.000								
Valor Mercado activo fijo									
Valor Mercado capital de Trabajo									200.000
Préstamo									
Amortización préstamo									
Flujo de Caja de Capitales	-7.085.000	-	-	-	-	-	-	-	200.000
Flujo Caja Privado	-7.085.000	1.472.562	1.938.716	1.943.378	1.948.087	1.952.844	3.831.410	3.836.262	2.167.399

7.3. Análisis de Sensibilidad

Para estar completamente seguros de que el proyecto es rentable, se realizó una sensibilización del supuesto de la tasa de renovación de contratos atribuible al proyecto (Tabla 7.6). Se determinó que con un 3 % de renovaciones logradas, equivalente a 6 contratos de mínimo 3 meses cada uno, el proyecto tendría un VAN positivo. Se estima que se puede conseguir sin problemas con la implementación del proyecto.

Tabla 7.6: Sensibilización del supuesto de renovaciones de contrato atribuibles al proyecto.

Tasa Renovación (%)	Incremento Clientes	VAN (CLP \$)	TIR (%)
1	2	-4.099.312	-10
2	3	-3.142.656	-3
3	6	246.221	11
4	8	2.274.514	18
5	10	4.507.259	25
10	17	11.528.966	40

Capítulo 8

Resultados

La presente sección tiene por objetivo analizar los datos recolectados para aceptar o rechazar las hipótesis planteadas anteriormente. El análisis se realizó utilizando lenguaje R en la IDE RStudio (RStudio Team, 2015).

8.1. Extracción, transformación y limpieza de datos

Considerando el siguiente conjunto de criterios, se seleccionaron 23.282 recomendaciones de las 1.139.193 emitidas hasta mediados de Enero 2017. Los criterios de selección fueron los siguientes:

1. Recomendaciones emitidas entre Septiembre y Diciembre 2016, dado que muchos anuncios utilizados vencían en Enero 2017.
2. Se excluyen todas las recomendaciones que incluyeron un anuncio de la “Bolsa de Trabajo”. Este anuncio suele tener una alta tasa de clics, por lo que fue eliminado para evitar ruido.

De acuerdo al diseño experimental, las recomendaciones con Formato tipo “texto” tenían 6 anuncios, mientras que las de tipo “imagen” sólo dos anuncios. Se observó en el conjunto de datos que algunas recomendaciones presentaron más de 2 clics, por lo tanto, para evitar un potencial desbalance de clics totales, se considerará que el máximo de clics por recomendación es igual a 2. Adicionalmente, extrae el número de la semana del año de la fecha en la que se produjo la recomendación.

El proceso descrito anteriormente ejecutó utilizando el código que se muestra en Código 8.1 y se obtiene los datos como se muestra de la línea 24 en adelante. Cada recomendación tiene un ID único, una marca de tiempo (*timestamp*), una corrida del experimento asociada, los clics obtenidos y la semana del año correspondiente.

```
1 > df.recs <- m.recs$aggregate('[
2   {"$match": {"$or": [{ "piwik_box_load_logged": true },
3   {"recommendations.clicks": {"$gt": 0} } ]
```



```

4         ] } },
5     {"$match": {"recommendations.target": {"$not": {"$regex": ".*trabajando.*"}}}},
6     {"$match": {"recommendations.category": {"$ne": "Agrotecnia"}}},
7     {"$project": {"timestamp": 1, "recommendations": 1, "experiment_run": 1 }},
8     {"$unwind": "$recommendations" },
9     {"$group": {
10      "_id": "$_id",
11      "timestamp": {"$first": "$timestamp"},
12      "experiment_run": {"$first": "$experiment_run"},
13      "clicks": {"$sum": {"$cond": [{"$gte": ["$recommendations.clicks", 1]}, 1, 0 ]}}
14    }
15  ]}')
16
17 > data <- df.recs %>%
18   filter(timestamp >= "2016-09-01" & timestamp < "2017-01-01") %>%
19   mutate(week = as.numeric(format(timestamp, "%U")),
20          clicks = if_else(clicks > 2, 2, as.double(clicks)))
21
22 > head(data)
23 ##           _id           timestamp experiment_run clicks week
24 ## 1 5838336b0b905f050d1923fa 2016-11-25 06:49:47           7     0    47
25 ## 2 583832c20b905f050d1923eb 2016-11-25 06:46:58           8     0    47
26 ## 3 583831140b905f050d1923d8 2016-11-25 06:39:48           5     0    47
27 ## 4 5808c8170b905f4d9040fed8 2016-10-20 07:35:19           4     0    42
28 ## 5 57dfbfb10b905f097a2c50a4 2016-09-19 04:36:33           8     0    38
29 ## 6 583830e40b905f050d1923d3 2016-11-25 06:39:00           8     0    47

```

Código 8.1: Importación, filtrado y primera transformación de los datos.

Explorando la distribución de las recomendaciones realizadas por cada corrida del experimento cada semana (Código 8.2) notamos que las semanas 40, 41 y 42 no tienen recomendaciones en las corridas 1, 3, 5 y 7. Además las semanas 43 y 44 tiene menos de 10 recomendaciones en las corridas 3 y 7 (línea 9-13). Esto ocurrió porque al poblar unos anuncios nuevos en la base de datos, se omitió un campo obligatorio para el despliegue de los anuncios y no se percibió del error en ese momento, dado que el sistema inyectaba los anuncios de manera asíncrona y simplemente no se desplegaba el bloque `div` en el sitio para esas corridas.

```

1 > table(data$week, data$experiment_run)
2 ##
3 ##      1  2  3  4  5  6  7  8
4 ## 35  71 130 23  75  70 136 34  76
5 ## 36 164 301 82 198 161 292 78 219
6 ## 37 165 330 99 210 148 331 100 234
7 ## 38 110 248 75 191 114 235 87 195
8 ## 39 93 218 39 192 88 217 51 181
9 ## 40 0 178 0 169 0 168 0 178
10 ## 41 0 104 0 143 0 108 0 133
11 ## 42 0 128 0 151 0 131 0 128
12 ## 43 112 168 5 139 117 148 8 132
13 ## 44 96 52 6 56 99 65 6 49
14 ## 45 209 66 46 105 230 72 55 89
15 ## 46 403 76 64 67 400 80 73 79
16 ## 47 411 40 184 78 421 49 175 106
17 ## 48 460 52 304 123 472 52 326 127
18 ## 49 495 81 359 81 507 73 343 80
19 ## 50 521 51 417 110 530 61 428 127
20 ## 51 540 63 421 137 506 52 459 165
21 ## 52 509 59 410 106 482 67 403 107

```

Código 8.2: Distribución de recomendaciones por corrida y semana del año.

Dado que éstas semanas con problemas técnicos representan sólo el 12,8% de los clics de todo el periodo (Código 8.3, línea 7-11), se decide eliminarlas (línea 13). Finalmente quedamos con 20.305 observaciones con la misma estructura de datos detallada anteriormente.

```

1 > data %>%
2 mutate(grupo = if_else(week %in% c(40,41,42,43,44), "Semana 40-44", "Resto")) %>%
3 group_by(grupo) %>%
4 summarize(clicks = sum(clicks), count = n()) %>%
5 mutate(freq.clicks = clicks / sum(clicks), freq.count = count / sum(count))
6
7 ## # A tibble: 2 x 5
8 ##   grupo clicks count freq.clicks freq.count
9 ##   <chr> <dbl> <int> <dbl> <dbl>
10 ## 1 Resto 995 20305 0.94045369 0.872133
11 ## 2 Semana 40-44 63 2977 0.05954631 0.127867
12
13 > data <- data %>% filter(!week %in% c(40,41,42,43,44))
14 > head(data)
15 ##   _id timestamp experiment_run clicks week
16 ## 1 5838336b0b905f050d1923fa 2016-11-25 06:49:47 7 0 47
17 ## 2 583832c20b905f050d1923eb 2016-11-25 06:46:58 8 0 47
18 ## 3 583831140b905f050d1923d8 2016-11-25 06:39:48 5 0 47
19 ## 4 57dfbfb10b905f097a2c50a4 2016-09-19 04:36:33 8 0 38
20 ## 5 583830e40b905f050d1923d3 2016-11-25 06:39:00 8 0 47
21 ## 6 585e25060b905f1e0a3df2de 2016-12-24 01:34:29 3 0 51

```

Código 8.3: Eliminación de semanas sin datos para algunas corridas del experimento.

El siguiente paso es agregar las observaciones por semana y corrida del experimento de acuerdo al diseño experimental y calculando el CTR. Obtenemos un set de datos con la estructura que se muestra en el Código 8.4.

```

1 > data.we <- data %>%
2 group_by(.dots = c("week", "experiment_run")) %>%
3 summarize(obs = n(),
4           total.clicks = sum(clicks),
5           ctr = (total.clicks/obs)*100)
6
7 > head(data.we)
8 ## Source: local data frame [6 x 5]
9 ## Groups: week [1]
10 ##
11 ##   week experiment_run obs total.clicks ctr
12 ##   <dbl> <int> <int> <dbl> <dbl>
13 ## 1 35 1 71 11 15.492958
14 ## 2 35 2 130 8 6.153846
15 ## 3 35 3 23 5 21.739130
16 ## 4 35 4 75 1 1.333333
17 ## 5 35 5 70 6 8.571429
18 ## 6 35 6 136 6 4.411765

```

Código 8.4: Eliminación de semanas sin datos para algunas corridas del experimento.

Cabe mencionar que también el diseño experimental establecía que los datos se analizarían de acuerdo al País y el dispositivo, pero no resultó posible por dos razones: (1) el autor renunció a su trabajo y perdió el acceso a los logs originales, y (2) los datos parciales que se mantuvieron no contenían esta información para el 50,7% de las recomendaciones, por lo que se trabajará con los datos a nivel de semana únicamente en este análisis.

Graficando la distribución del conjunto de datos con la estructura final para el análisis observamos que existen potenciales *outliers* (Figura 8.1). Particularmente nos interesa determinar si hay *outliers* que eliminar en la variable CTR.

Si observamos los potenciales *outliers* (Código 8.5), podemos ver que los hay dos relacionados a la corrida 3 en dos semanas diferentes. El número de observaciones obtenidas es casi tres veces superior a los clics obtenidos, por lo que se infiere que son semanas de buena

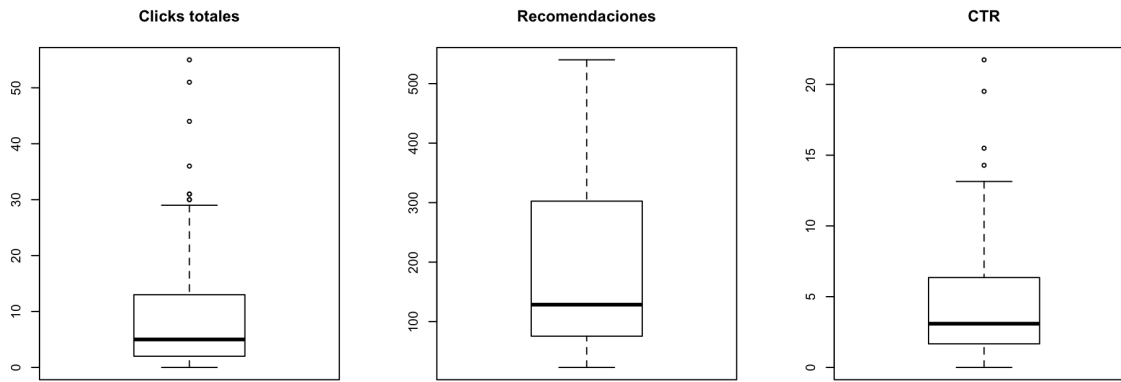


Figura 8.1: Distribución de variables del conjunto de datos.

conversión y no debería ser descartadas como outliers.

```

1 > filter(data.we, ctr %in% bp.ctr$out)
2 ## Source: local data frame [4 x 5]
3 ## Groups: week [3]
4 ##
5 ##   week experiment_run  obs total.clicks      ctr
6 ##   <dbl>          <int> <int>      <dbl>    <dbl>
7 ## 1     35             1    71         11 15.49296
8 ## 2     35             3    23          5 21.73913
9 ## 3     36             3    82         16 19.51220
10 ## 4     51             2    63          9 14.28571

```

Código 8.5: Potenciales *outliers*.

Para determinar los dos casos restantes, correspondientes a la corrida 1 y 2, se observa en los resultados (Código 8.6) que las conversiones más altas de los dos subconjuntos de datos anteriores (corrida 1 = 15,5%; y corrida 2 = 14,3%) son de valores cercanos entre sí, con un número de observaciones similares para algunas de las semanas restantes.

El valor máximo para la corrida 1 se obtuvo con 71 observaciones, ubicadas en el percentil 7,7%, durante la primera semana del experimento. Dado que es probable que el efecto de la novedad produjo este alto valor, este valor se mantendrá en el set de datos porque provee información relevante. Más adelante se analizará si afecta la normalidad de los datos (en términos de distribución estadística).

Para el caso de la corrida 2, la conversión máxima se obtuvo con 63 observaciones, pero en este caso, éstas se encuentra en el percentil 38,5% de los datos. Por las mismas razones señaladas anteriormente para la corrida 1, se mantendrá este dato.

```

1 > data.we %>% filter(experiment_run == 1) %>% arrange(obs) -> corrida1
2 > corrida1$percentile <- ecdf(corrida1$obs)(corrida1$obs)
3 > corrida1
4 ## Source: local data frame [13 x 6]
5 ## Groups: week [13]
6 ##
7 ##   week experiment_run  obs total.clicks      ctr percentile
8 ##   <dbl>          <int> <int>      <dbl>    <dbl>    <dbl>
9 ## 1     35             1    71         11 15.492958 0.07692308
10 ## 2     39             1    93          6  6.451613 0.15384615
11 ## 3     38             1   110          4  3.636364 0.23076923

```

```

12 ## 4      36          1    164          13  7.926829  0.30769231
13 ## 5      37          1    165          5  3.030303  0.38461538
14 ## 6      45          1    209          5  2.392344  0.46153846
15 ## 7      46          1    403          20  4.962779  0.53846154
16 ## 8      47          1    411          10  2.433090  0.61538462
17 ## 9      48          1    460          20  4.347826  0.69230769
18 ## 10     49          1    495          17  3.434343  0.76923077
19 ## 11     52          1    509          30  5.893910  0.84615385
20 ## 12     50          1    521          31  5.950096  0.92307692
21 ## 13     51          1    540          51  9.444444  1.00000000
22
23 > data.we %>% filter(experiment_run == 2) %>% arrange(obs) -> corrida2
24 > corrida2$percentile <- ecdf(corrida2$obs)(corrida2$obs)
25 > corrida2
26 ## Source: local data frame [13 x 6]
27 ## Groups: week [13]
28 ##
29 ##   week experiment_run  obs total.clicks      ctr percentile
30 ##   <dbl>          <int> <int>      <dbl>    <dbl>    <dbl>
31 ## 1      47            2    40         0  0.000000  0.07692308
32 ## 2      50            2    51         0  0.000000  0.15384615
33 ## 3      48            2    52         0  0.000000  0.23076923
34 ## 4      52            2    59         1  1.694915  0.30769231
35 ## 5      51            2    63         9 14.285714  0.38461538
36 ## 6      45            2    66         0  0.000000  0.46153846
37 ## 7      46            2    76         1  1.315789  0.53846154
38 ## 8      49            2    81         0  0.000000  0.61538462
39 ## 9      35            2   130        8  6.153846  0.69230769
40 ## 10     39            2   218         6  2.752294  0.76923077
41 ## 11     38            2   248         7  2.822581  0.84615385
42 ## 12     36            2   301         9  2.990033  0.92307692
43 ## 13     37            2   330        13  3.939394  1.00000000

```

Código 8.6: Exploración de datos relacionados a las corridas 1 y 2.

Graficando la distribución de los valores CTR para cada semana y corrida (Figura 8.2), notamos que hay corridas que son más eficientes que otras, en particular la corrida 3 y 7. Lo interesante que nos muestra esta gráfica es que las corridas 3 y 7 son las que consideran anuncios de tipo texto con una relevancia semántica alta, lo que estaría validando la Hipótesis 1. Adicionalmente se observa una baja en el CTR entre las semanas 38, 39 y 45, las que justamente coinciden con el inicio de temporada de varios cultivos del hemisferio sur, momentos en que los productores probablemente tengan menos tiempo para "leer noticias".

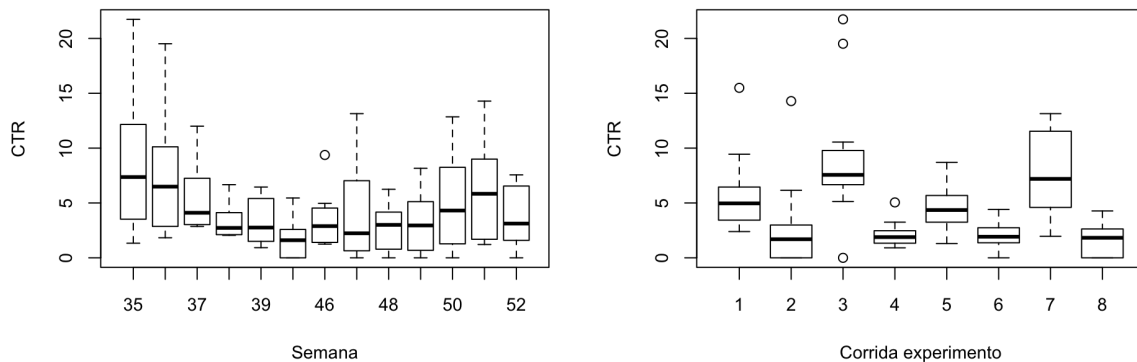


Figura 8.2: Distribución de variables del conjunto de datos.

Si obtenemos el promedio de CTR por corrida notamos que todas corridas obtuvieron un CTR mayor a la conversión del sistema tradicional de anuncios de la empresa (0,24%) o al promedio mundial para banners reportado por Google (0,35%) o al promedio de conversión de anuncios en Google Adwords (7%).

```

1 > data %>%
2   group_by(experiment_run) %>%
3   summarize(total.clicks = sum(clicks),
4             obs = n(),
5             ctr = (total.clicks / obs) * 100 )
6 ## # A tibble: 8 x 4
7 ##   experiment_run total.clicks  obs      ctr
8 ##   <int>          <dbl> <int>  <dbl>
9 ## 1             1           223  4151  5.372199
10 ## 2             2            54  1715  3.148688
11 ## 3             3           213  2523  8.442331
12 ## 4             4            38  1673  2.271369
13 ## 5             5           180  4129  4.359409
14 ## 6             6            39  1717  2.271404
15 ## 7             7           213  2612  8.154671
16 ## 8             8            35  1785  1.960784

```

Código 8.7: Promedios CTR por corrida.

Finalmente, antes de entrar al análisis estadístico, debemos traducir el número de las corridas en factores y sus niveles para poder ser analizadas en R (Código 8.8).

```

1 > data.agg <- data.we %>%
2   mutate(formato = if_else(experiment_run %in% c(2,4,6,8), "Imagen", "Texto"),
3          relevancia = if_else(experiment_run %in% c(3,4,7,8), "Alta", "Baja"),
4          etiquetado = if_else(experiment_run %in% c(5,6,7,8), "Si", "No")
5   )
6 > head(data.agg)
7 ## Source: local data frame [6 x 8]
8 ## Groups: week [1]
9 ##
10 ##   week experiment_run  obs total.clicks      ctr formato relevancia
11 ##   <dbl>          <int> <int>      <dbl>    <dbl> <chr>    <chr>
12 ## 1     35             1     71      11 15.492958  Texto    Baja
13 ## 2     35             2    130       8  6.153846  Imagen   Baja
14 ## 3     35             3     23       5 21.739130  Texto    Alta
15 ## 4     35             4     75       1  1.333333  Imagen   Alta
16 ## 5     35             5     70       6  8.571429  Texto    Baja
17 ## 6     35             6    136       6  4.411765  Imagen   Baja
18 ## # ... with 1 more variables: etiquetado <chr>
19
20 > data.agg$formato <- as.factor(data.agg$formato)
21 > data.agg$relevancia <- as.factor(data.agg$relevancia)
22 > data.agg$etiquetado <- as.factor(data.agg$etiquetado)
23 > summary(data.agg)
24 ##      week      experiment_run      obs      total.clicks
25 ##   Min.   :35.00   Min.   :1.00   Min.   : 23.00   Min.   : 0.000
26 ##   1st Qu.:38.00   1st Qu.:2.75   1st Qu.: 75.75   1st Qu.: 2.000
27 ##   Median :46.00   Median :4.50   Median :128.50   Median : 5.000
28 ##   Mean   :44.08   Mean   :4.50   Mean   :195.24   Mean   : 9.567
29 ##   3rd Qu.:49.00   3rd Qu.:6.25   3rd Qu.:301.75   3rd Qu.:13.000
30 ##   Max.   :52.00   Max.   :8.00   Max.   :540.00   Max.   :55.000
31 ##      ctr      formato  relevancia  etiquetado
32 ##   Min.   : 0.000   Imagen:52  Alta:52    No:52
33 ##   1st Qu.: 1.681   Texto :52  Baja:52    Si:52
34 ##   Median : 3.090
35 ##   Mean    : 4.504
36 ##   3rd Qu.: 6.300
37 ##   Max.    :21.739

```

Código 8.8: Configuración de factores y sus niveles del conjunto de datos.

8.2. Análisis estadístico

Para determinar el test estadístico a utilizar debemos comprobar si los datos tienen una distribución normal o no. Utilizando el test de Shapiro (Código 8.9) se observa que las corridas 1 a 4 tienen una distribución no normal ($p\text{-value} < 0,05$). Las primeras cuatro corridas corresponde a los que tenían el factor “Etiquetado” en nivel “Sin Etiquetar”.

```
1 > tapply(data.agg$ctr, data.agg$experiment_run, shapiro.test)
2 ## $'1'
3 ##
4 ## Shapiro-Wilk normality test
5 ##
6 ## data: X[[i]]
7 ## W = 0.8313, p-value = 0.01645
8 ##
9 ##
10 ## $'2'
11 ##
12 ## Shapiro-Wilk normality test
13 ##
14 ## data: X[[i]]
15 ## W = 0.71289, p-value = 0.0007491
16 ##
17 ##
18 ## $'3'
19 ##
20 ## Shapiro-Wilk normality test
21 ##
22 ## data: X[[i]]
23 ## W = 0.85279, p-value = 0.03099
24 ##
25 ##
26 ## $'4'
27 ##
28 ## Shapiro-Wilk normality test
29 ##
30 ## data: X[[i]]
31 ## W = 0.85118, p-value = 0.02953
32 ##
33 ##
34 ## $'5'
35 ##
36 ## Shapiro-Wilk normality test
37 ##
38 ## data: X[[i]]
39 ## W = 0.85406, p-value = 0.1047
40 ##
41 ##
42 ## $'6'
43 ##
44 ## Shapiro-Wilk normality test
45 ##
46 ## data: X[[i]]
47 ## W = 0.97028, p-value = 0.8976
48 ##
49 ##
50 ## $'7'
51 ##
52 ## Shapiro-Wilk normality test
53 ##
54 ## data: X[[i]]
55 ## W = 0.86099, p-value = 0.1228
56 ##
57 ##
58 ## $'8'
59 ##
60 ## Shapiro-Wilk normality test
61 ##
```

```
62 ## data: X[[i]]
63 ## W = 0.91493, p-value = 0.2142
```

Código 8.9: Aplicación del test de Shapiro por corrida.

Esto implica que deberán analizarse los resultados separadamente para las corridas 1-4, corridas 5-8 y un análisis exclusivo para el factor “Etiquetado”.

8.2.1. Análisis corridas 1 a 4

En primer lugar se debe normalizar la data, para ello se utiliza la función `rntransform` (Código 8.10). La transformación se comprueba utilizando el test Shapiro nuevamente, lo que refleja que los datos son muy normales ahora (p-value ≈ 1)

```
1 > data.agg.run1234 <- data.agg %>% filter(experiment_run < 5)
2 > data.agg.run1234$ctr.norm <- 0
3 > for(i in c(1,2,3,4)) {
4   ctr <- data.agg.run1234[data.agg.run1234$experiment_run == i, "ctr"]
5   data.agg.run1234[data.agg.run1234$experiment_run == i, "ctr.norm"] <- rntransform(ctr$ctr)
6 }
7 > tapply(data.agg.run1234$ctr.norm, data.agg.run1234$experiment_run, shapiro.test)
8 ## $'1'
9 ##
10 ## Shapiro-Wilk normality test
11 ##
12 ## data: X[[i]]
13 ## W = 0.99811, p-value = 1
14 ##
15 ##
16 ## $'2'
17 ##
18 ## Shapiro-Wilk normality test
19 ##
20 ## data: X[[i]]
21 ## W = 0.88839, p-value = 0.09274
22 ##
23 ##
24 ## $'3'
25 ##
26 ## Shapiro-Wilk normality test
27 ##
28 ## data: X[[i]]
29 ## W = 0.99811, p-value = 1
30 ##
31 ##
32 ## $'4'
33 ##
34 ## Shapiro-Wilk normality test
35 ##
36 ## data: X[[i]]
37 ## W = 0.99811, p-value = 1
```

Código 8.10: Aplicación de `rntransform` para normalizar los datos.

El Código 8.11 muestra los resultados del test de Anova, del que se desprende:

- El promedio de CTR obtenido por las diferentes combinaciones del experimento fue de 4,95 %, superior al promedio de conversión pre-experimento igual a 0,24 % del sitio web y el 0,35 % de conversión promedio reportado para banners por Google. Sin embargo fue inferior al 7 % de conversión promedio reportado para Google Adwords.

- El promedio de CTR para el factor “Formato” fue de 2,4% para “Imagen” y 7,5% para “Texto”, diferencia significativa según la tabla Anova ($p = 3,18e-05$). Importante destacar que este promedio supera la conversión del 7% reportada por Google Adwords como promedio mundial.
- El resultado de la línea 24 muestra que el factor formato fue el más significativo (p -value = $3,18e-05$), mientras que la relevancia por sí misma no produjo una diferencia significativa ($p = 0,23$), y la interacción entre Formato:Relevancia no fue significativa en todas sus combinaciones. Esto estaría validando la Hipótesis 2 de que los anuncios tipo texto son más efectivos que los de imagen en este caso.
- Utilizando el test TukeyHSD (Línea 33) vemos que cuando la relevancia se combina con anuncios de texto, el CTR se ve significativamente aumentado y superior a los anuncios de imagen, independiente que éstos tengan relevancia alta o baja. Por lo tanto, se estaría rechazando la Hipótesis 1 de que la relevancia por si sola produciría un aumento en la conversión.

```

1 > aov.out.run1234 <- aov(with(data.agg.run1234, lm(ctr~formato*relevancia)))
2 > model.tables(aov.out.run1234, type = "mean")
3 ## Tables of means
4 ## Grand mean
5 ##
6 ## 4.947269
7 ##
8 ## formato
9 ## formato
10 ## Imagen Texto
11 ## 2.423 7.472
12 ##
13 ## relevancia
14 ## relevancia
15 ## Alta Baja
16 ## 5.612 4.283
17 ##
18 ## formato:relevancia
19 ##      relevancia
20 ## formato Alta Baja
21 ## Imagen 2.079 2.766
22 ## Texto 9.144 5.800
23
24 > summary(aov.out.run1234)
25 ##              Df Sum Sq Mean Sq F value    Pr(>F)
26 ## formato      1  331.4    331.4  21.087 3.18e-05 ***
27 ## relevancia   1   23.0     23.0   1.461 0.233
28 ## formato:relevancia 1   52.8     52.8   3.359 0.073 .
29 ## Residuals   48  754.5     15.7
30 ## ---
31 ## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
32
33 > TukeyHSD(aov.out.run1234, ordered = TRUE)
34 ## Tukey multiple comparisons of means
35 ## 95% family-wise confidence level
36 ## factor levels have been ordered
37 ##
38 ## Fit: aov(formula = with(data.agg.run1234, lm(ctr ~ formato * relevancia)))
39 ##
40 ## $formato
41 ##          diff      lwr      upr      p adj
42 ## Texto-Imagen 5.049329 2.838474 7.260183 3.18e-05
43 ##
44 ## $relevancia
45 ##          diff      lwr      upr      p adj
46 ## Alta-Baja 1.32904 -0.8818151 3.539894 0.2327069
47 ##
48 ## $'formato:relevancia'
49 ##          diff      lwr      upr      p adj

```



```

50 ## Imagen:Baja-Imagen:Alta 0.6862632 -3.4522817 4.824808 0.9709675
51 ## Texto:Baja-Imagen:Alta 3.7202888 -0.4182561 7.858834 0.0922066
52 ## Texto:Alta-Imagen:Alta 7.0646314 2.9260865 11.203176 0.0002139 |\label{line:tukey-texto-image-alta}|
53 ## Texto:Baja-Imagen:Baja 3.0340256 -1.1045193 7.172571 0.2209344
54 ## Texto:Alta-Imagen:Baja 6.3783682 2.2398233 10.516913 0.0008820
55 ## Texto:Alta-Texto:Baja 3.3443426 -0.7942024 7.482887 0.1520342 |\label{line:tukey-texto-relevancia}|

```

Código 8.11: Aplicación del test Anova para las corridas 1 a 4.

8.2.2. Análisis corridas 5 a 8

Dado que las corridas 5 a 8 presentaron una distribución normal, podemos aplicar el test de Anova directamente (Código 8.12) del que se desprende:

- El promedio de CTR para éstas corridas fue de 3,5 %, menor que las corridas 1-4.
- Al igual que las corridas 1-4 el factor formato también produce una diferencia significativa de 4,6 %, siendo 1,8 % el promedio para imágenes y 6,4 % para texto. Esto estaría validando la Hipótesis 2 de que los anuncios tipo texto son más efectivos que los de imagen en este caso.
- A diferencia del otro grupo de corridas, todas las interacciones entre formato y relevancia fueron significativas excepto cuando se trató de sólo imágenes y sólo texto (Línea 36). Por lo tanto, se estaría rechazando la Hipótesis 1 de que la relevancia por si sola produciría un aumento en la conversión.

```

1 > data.agg.run5678 <- data.agg %>% filter( experiment_run >= 5 )
2 > aov.out.run5678 <- aov(with(data.agg.run5678, lm(ctr~formato*relevancia)))
3 > model.tables(aov.out.run5678, type = "mean")
4 ## Tables of means
5 ## Grand mean
6 ##
7 ## 3.569586
8 ##
9 ## formato
10 ##      Imagen  Texto
11 ##      1.836  6.387
12 ## rep 26.000 16.000
13 ##
14 ## relevancia
15 ##      Alta  Baja
16 ##      3.896  3.244
17 ## rep 21.000 21.000
18 ##
19 ## formato:relevancia
20 ##      relevancia
21 ## formato Alta  Baja
22 ## Imagen  1.693  1.978
23 ## rep    13.000 13.000
24 ## Texto   7.474  5.300
25 ## rep     8.000  8.000
26
27 > summary(aov.out.run5678)
28 ##              Df Sum Sq Mean Sq F value    Pr(>F)
29 ## formato          1  205.18   205.18  32.444 1.49e-06 ***
30 ## relevancia        1    4.46    4.46   0.706  0.406
31 ## formato:relevancia 1   14.98   14.98   2.368  0.132
32 ## Residuals       38  240.31    6.32
33 ## ---
34 ## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
35
36 > TukeyHSD(aov.out.run5678, ordered = TRUE)

```

```

37 ## Tukey multiple comparisons of means
38 ## 95% family-wise confidence level
39 ## factor levels have been ordered
40 ##
41 ## Fit: aov(formula = with(data.agg.run5678, lm(ctr ~ formato * relevancia)))
42 ##
43 ## $formato
44 ##          diff          lwr          upr          p adj
45 ## Texto-Imagen 4.551388 2.933787 6.168988 1.5e-06
46 ##
47 ## $relevancia
48 ##          diff          lwr          upr          p adj
49 ## Alta-Baja 0.652101 -0.9189805 2.223182 0.4060212
50 ##
51 ## $'formato:relevancia'
52 ##          diff          lwr          upr          p adj
53 ## Imagen:Baja-Imagen:Alta 0.2847628 -2.3651064 2.934632 0.9914793
54 ## Texto:Baja-Imagen:Alta 3.6065166 0.5707101 6.642323 0.0144434
55 ## Texto:Alta-Imagen:Alta 5.7810213 2.7452148 8.816828 0.0000531
56 ## Texto:Baja-Imagen:Baja 3.3217538 0.2859472 6.357560 0.0273252
57 ## Texto:Alta-Imagen:Baja 5.4962585 2.4604520 8.532065 0.0001157
58 ## Texto:Alta-Texto:Baja 2.1745047 -1.2034289 5.552438 0.3230203

```

Código 8.12: Aplicación del test Anova para las corridas 5 a 8.

8.2.3. Análisis para el factor “Etiquetado”

Dado que los resultados del experimento indican que el conjunto de observaciones con el factor “Etiquetado” tienen una distribución no normal para el nivel positivo (Etiquetado “Sin Etiquetar”) y para el resto, el comportamiento es de tipo normal, los resultados del análisis Anova previamente descritos no son comparables.

Por lo tanto, desconocemos la incidencia de este factor sobre la variable objetivo (CTR). Como análisis complementario se determinará si existe una correlación entre el etiquetado y la variable objetivo. Para ello debemos conocer si el conjunto de datos presenta Normalidad al separar las observaciones de acuerdo al nivel del Factor. Al aplicar el test de Shapiro, se determina que los datos presentan una distribución normal (Código 8.13)

```

1 > tapply(data.agg$ctr, data.agg$etiquetado, shapiro.test)
2 ## $No
3 ##
4 ## Shapiro-Wilk normality test
5 ##
6 ## data: X[[i]]
7 ## W = 0.83587, p-value = 4.611e-06
8 ##
9 ##
10 ## $Si
11 ##
12 ## Shapiro-Wilk normality test
13 ##
14 ## data: X[[i]]
15 ## W = 0.84259, p-value = 4.11e-05

```

Código 8.13: Aplicación del test Shapiro para el Factor “Etiquetado”.

Dado que el conjunto de datos presenta una distribución normal, podemos determinar la significancia del efecto utilizando el test T-Student (Código 8.14)

```

1 > data.etiquetado <- data.agg[order(data.agg$etiquetado, data.agg$week),]

```

```

2 > t.test(data.etiquetado$ctr ~ data.etiquetado$etiquetado)
3 ##
4 ## Welch Two Sample t-test
5 ##
6 ## data: data.etiquetado$ctr by data.etiquetado$etiquetado
7 ## t = 1.6373, df = 90.485, p-value = 0.105
8 ## alternative hypothesis: true difference in means is not equal to 0
9 ## 95 percent confidence interval:
10 ## -0.2938756 3.0492404
11 ## sample estimates:
12 ## mean in group No mean in group Si
13 ## 4.947269 3.569586

```

Código 8.14: Aplicación del test T-Student para el Factor “Etiquetado”.

De los resultados anteriores se desprende que:

- De acuerdo a los resultados del test T-Student, la diferencia media de CTR obtenida el grupo con observaciones etiquetadas y no etiquetadas no es significativa (p-value = 0.2772). Para las observaciones etiquetadas se obtuvo un promedio de 4,06 %, mientras para las observaciones no etiquetadas se obtuvo un promedio de 4,95 %.
- Se rechaza la Hipótesis 3 dado que no hay diferencia significativa entre los niveles del factor.

8.2.4. Clics totales

Si se comparan los clics registrados en Piwik durante Septiembre-Diciembre, se concluye que el nuevo espacio publicitario demostró contribuir significativamente a la producción de clics, ya que generó 2.522 clics (en 1 espacio publicitario), equivalente al 29,3 % de los 8.596 clics totales generados por el resto de las 19 posiciones publicitarias del sitio web durante el mismo periodo.

Lo anterior valida la Hipótesis 4 de que un espacio publicitario dinámico es más efectivo que los estáticos, incluso, podría ser más efectivo si solo se considera utilizar la mejor combinación de factores.

Capítulo 9

Discusión

Al momento de realizar el análisis del presente trabajo, el autor ya no trabajaba en la empresa en la que se implementó el experimento, por lo que hay ciertos análisis de contexto útiles que ya no pueden ser realizados. Por ejemplo, sería interesante observar la distribución de visitas del sitio web durante el periodo del experimento.

De haber tenido dicho dato, podrían haberse estandarizado las visitas por semana, como también analizar si la recurrencia presenta una diferencia significativa en la conversión con este sistema de recomendación. Esto afectó particularmente la respuesta a la Interrogante 4 que pretendía encontrar si existía una diferencia estadística en los usuarios nuevos y recurrentes, por lo tanto, tampoco pudo ser comprobado el efecto de novedad. Como aprendizaje, la recurrencia debería haberse almacenado directamente con cada recomendación.

En la misma línea, otro aprendizaje es implementar *Unit testing* para verificar que todos los aspectos del software estén funcionando correctamente y no confiarse en sólo algunas verificaciones. Evidentemente esto retrasa el desarrollo cuando no se dispone de un *Quality assurance (QA)*, pero puede evitar que se pierdan 4 semanas de datos como ocurrió en este trabajo.

En la Figura 8.2 “Distribución de variables del conjunto de datos” se observó un potencial patrón de comportamiento clásico de exportadores/productores chilenos, pero éstos datos corresponden a las recomendaciones registradas y no a las visitas totales del sitio web, datos a los que ya no se puede acceder para realizar una test estadístico exploratorio. Considerando lo anterior, algunos anuncios publicitarios debieran ser más efectivos en ciertas épocas del año y lo que debería incluirse en la priorización de las recomendaciones. Esto sin duda será agregado como Trabajo Futuro.

El Factor “Relevancia” no produjo un incremento significativo en la conversión en este experimento por sí sola, pero si se aprecian incrementos significativos cuando interacciona, particularmente, con anuncios de texto. Por lo tanto la Hipótesis 1 “Objetivos específicos” se rechaza.

Sin embargo, este resultado contradice la experiencia la descrita en varios estudios ante-

riores (ver Sección 2.2.3). El autor estima que es posible, que dado la sobrecarga visual del sitio web, haya sido significativamente más relevante para los lectores “no ver banners” en lugar de atender al contenido de los mismos. Quizás los lectores de Portalfruticola.com ya estén completamente acostumbrados a ignorar imágenes fuera del espacio de la noticia porque el layout del sitio web no ha variado desde su creación (válido para lectores recurrentes) y porque es el mismo layout que se utiliza en varios sitios web (válido para los lectores nuevos).

Muy en relacionado con lo anterior, el Factor “Formato” demostró producir diferencias en la conversión significativa, tanto para el grupo de las corridas 1 a 4 y 5 a 8. Esto valida la Hipótesis 2 “Objetivos específicos” y refuerza sospecha que la sobrecarga del sitio web esté afectando la efectividad de su servicio central.

Por el lado del Factor “Etiquetado”, este no produjo una diferencia significativa en la conversión, pero dado que el grupo de las corridas 1 a 4 (sin etiquetado) presentó una conversión promedio superior (5,0%) a las corridas 5 a 8 (con etiquetado, promedio CTR = 3,5%) y dada la sospecha de la saturación gráfica del sitio web, se recomienda no etiquetar los anuncios publicitarios. Éste resultado rechazaría la Hipótesis 3 “Objetivos específicos”.

Particularmente interesante es lo efectivo que demostró ser este nuevo espacio publicitario en relación a los espacios tradicionales, ya que los clics capturados equivalen al 29,3% de los clics convertidos por el resto de las 19 posiciones de anuncios publicitarios del sitio. De potenciarse esta sección y utilizar la mejor combinación de factores, podrían eliminarse las posiciones de anuncios menos valoradas por los clientes y aún así hacer más efectivas las campañas. Esto valida la Hipótesis 4 “Objetivos específicos”.

Avanzando hacia la elección de combinación de factores óptima, se considera que la elección del formato de texto y el no etiquetado es evidente en base al análisis estadístico y argumento expuestos. Pese a que el factor de relevancia semántica no fue por si sólo significativo, si lo es en su interacción con el formato.

Si vemos en detalle el los resultados de interacción entre el factor Formato y Relevancia del conjunto de datos con el factor Etiquetado en su nivel bajo (ausente) en Código 9.1, encontramos que existe una diferencia significativa entre los anuncios de formato texto e imagen cuando ambos tienen una relevancia alta (Línea 5). El máximo CTR obtenido fue de 11,2% siendo la diferencia con los anuncios de imágenes relevantes de 7% si los comparamos con los de texto relevantes.

Para el mismo bloque de Código, cuando comparamos el efecto de la relevancia entre anuncios de texto (Línea 8), vemos que la diferencia no fue significativa, pero la diferencia fue de 3% y el valor máximo fue de 7,5%.

```

1 ## $'formato:relevancia'
2 ##
3 ## Imagen:Baja-Imagen:Alta 0.6862632 -3.4522817 4.824808 0.9709675
4 ## Texto:Baja-Imagen:Alta 3.7202888 -0.4182561 7.858834 0.0922066
5 ## Texto:Alta-Imagen:Alta 7.0646314 2.9260865 11.203176 0.0002139
6 ## Texto:Baja-Imagen:Baja 3.0340256 -1.1045193 7.172571 0.2209344
7 ## Texto:Alta-Imagen:Baja 6.3783682 2.2398233 10.516913 0.0008820
8 ## Texto:Alta-Texto:Baja 3.3443426 -0.7942024 7.482887 0.1520342

```

Código 9.1: Parte del test Tukey HSD para las corridas 1 a 4.

Considerando lo anterior, el autor recomienda que las dos mejores combinaciones de factores y sus niveles son:

1. Anuncios de texto semánticamente relevantes sin etiquetado.
2. Anuncios de texto semánticamente no relevantes sin etiquetado.

El uso de anuncios de texto en este nuevo espacio publicitario es evidente. Sin embargo la propuesta de utilizar o no el sistema de recomendación semántico, sólo se debe a la mantención del segundo es más sencilla, pero sólo la utilización de anuncios relevantes asegura el tener una conversión interesante (Cuadro 9.1).

Tabla 9.1: CTR promedio de anuncios de texto con diferentes relevancias

Corrida	Relevancia Alta	Relevancia Baja
Corrida 1-4	9,1	5,8
Corrida 5-8	7,5	5,3

Desde el punto de vista del negocio, **se puede afirmar que este nuevo espacio publicitario es competitivo en el mercado** frente a las grandes competidores de Portalfruticola.com, dado que se logró superar la conversión promedio mundial de Google AdSense de 0,35 % y de 7 % para Google Adwords, sin mencionar que supera con creces la conversión normal del sitio web de 0,24 %. El autor considera que esto son potentes de argumentos de venta para un sitio de nicho. Esto excede con creces el Resultado Esperado 1 “Resultados esperados”.

Cabe destacar que el sistema podría modificarse para que pueda soportar las lógicas modernas de la publicidad orientada a resultados y sus respectivos modelos de cobro como el CPM y CPC, por lo que el Resultado Esperado 3 “Resultados esperados” se considera cumplido. Algo que el nuevo de dueño de la empresa, Phoenix Media¹, demostró interés para algunos proyectos de sitios orientados a consumidor final.

Lamentablemente, el Resultado Esperado 2 “Resultados esperados” no pudo ser comprobado puesto que nunca se utilizaron los argumentos de ventas entregados por este trabajo ni se continuó en el desarrollo, mantención y mejoramiento de este sistema. De hecho la Dirección no se mostró interesada en mejorar el resultado de las campañas a sus clientes (ni la experiencia de lectura de sus lectores), pese que a clientes nacionales e internacionales dejaron de contratar el servicio por malos resultados. También se tiene antecedentes de que se han perdido renovaciones y potenciales clientes por no tener disponibles lógicas de cobro modernas como CPM y CPC.

Un aprendizaje sobre lo anterior, es que a veces, la agenda personal del Gerente General no es la misma que la agenda de la compañía.

¹Durante el transcurso de la Tesis, la empresa Yentzen Group fue comprada en su totalidad por la empresa de medios Phoenix Media

Capítulo 10

Trabajo futuro

En la discusión de los resultados se mencionó la sospecha de la sobrecarga visual del sitio web, pero esto no ha sido demostrado. Existen estudios de *Eye tracking* que permiten medir determinar efecto de elementos visuales en la atención de los usuarios, pero son estudios costos que serían difíciles de aplicar para la lectoría global de Portalfruticola.com. Sin duda se investigarán alternativas de poder resolver este punto, pero existe la opción de explorar el comportamiento de lectura utilizando herramientas web que literalmente graban lo que el usuario realiza en el sitio web y el movimiento del mouse. Otra opción sería contratar una asesoría experta en experiencia del usuario.

Desde el punto de vista tecnológico, para el registro de las visitas se utilizó Piwik porque permitía obtener un ID único para cada visita, característica que no era accesible en Google Analytics en el momento de implementación. Ahora que si lo está, lo ideal sería utilizar Google Analytics ya que proporciona más información de la visita que Piwik y se almacena en los servidores de Google. Dado que Piwik se almacena localmente, demandaba procesamiento al servidor.

Conectado con lo anterior, WordPress, que es el CMS de Portalfruticola.com, demanda significativos recursos de procesamiento dada la cantidad de plugin de terceros que se utilizan. Para mejorar la experiencia de lectura, e incluir sistemas como este de manera modular, se cree efectivo migrar a modelos modernos basados en API como Prismic.io.

La industria agrícola trabaja en base a temporadas, por lo que las personas deben tener comportamientos diferentes durante cada época del año. Este patrón de comportamiento sería muy interesante de agregar al sistema de recomendación para optimizar la priorización y despliegue de anuncios.

Muy en relación con lo anterior, se tiene la hipótesis de que los diferentes perfiles de la industria agrícola (ej: productores, exportadores, retailers, importadores y proveedores de servicios) debería tener un comportamiento de lectura e intereses diferentes. Existe algunas clasificaciones conocidas de perfiles para Portalfruticola.com (implementadas por el autor), por lo que podría analizarse si realmente “leen artículos diferentes”, en qué época del año leen y qué anuncios clickean. Si hay diferencias en los grupos, podría categorizarse el tráfico para

los lectores no registrados y optimizarse el despliegue de anuncios, contenido y búsqueda de clientes publicitarios.

Sobre lo anterior, se propuso trabajar después del presente trabajo, pero la Dirección no se mostró interesada.

Finalmente, dos puntos sobre los que puede optimizarse el sistema son los pesos del perfil y contenido para los ranking, como también la lógica para asignar vectores a anuncios publicitarios.

Capítulo 11

Conclusiones

El objetivo general del proyecto fue “Determinar los factores, su combinación y niveles, que aumentan los clics de las campañas publicitarias en Portalfruticola.com en un nuevo espacio publicitario durante el 2016”. Considerando los resultados, las dos mejores combinaciones de factores y sus niveles son:

1. Anuncios de texto semánticamente relevantes sin etiquetado.
2. Anuncios de texto semánticamente no relevantes sin etiquetado.

Finalmente, **se puede afirmar que este nuevo espacio publicitario es competitivo en el mercado** frente a los grandes competidores de Portalfruticola.com, dado que se logró superar la conversión promedio mundial de Google AdSense igual a 0,35% y de 7% para Google Adwords, sin mencionar que superó con creces la conversión normal del sitio web de 0,24%. El autor considera que esto son potentes de argumentos de venta para un sitio de nicho.

Desde el punto de vista del impacto en el negocio, este trabajo no produjo un cambio en el comportamiento de la empresa.

Capítulo 12

Bibliografía

- Acquisti, A., y Spiekermann, S. (2011). Do Interruptions Pay off? Effects of Interruptive Ads on Consumers' Willingness to Pay. *Journal of Interactive Marketing*, 25(4), 226–240. Descargado de <http://linkinghub.elsevier.com/retrieve/pii/S1094996811000326><http://dx.doi.org/10.1016/j.intmar.2011.04.003> doi: 10.1016/j.intmar.2011.04.003
- Asociación Chilena de Agencias de Publicidad (ACHAP). (2015). *Inversión publicitaria en medios*.
- Barros, O. (2009). *Ingeniería de Negocios: Diseño integrado de negocios, procesos y aplicaciones II*. Santiago de Chile: Universidad de Chile.
- Barros, O., y Julio, C. (2011). Enterprise and Process Architecture Patterns. *Business Process Management Journal*, 17(4), 598–618.
- Brady, M. P. (2009, oct). *Advertising effectiveness and spillover* (Vol. 25) (n.º 4). Descargado de <http://onlinelibrary.wiley.com/doi/10.1002/sdr.426/abstract><http://yb9vv9jh8j.search.serialssolutions.com?sid=CentralSearch:null{&}genre=article{&}atitle=Advertising+effectiveness+and+spillover{&}volume=25{&}issue=4{&}title=System+Dynamics+Review{&}issn=1099-1727{&}dat>
- Chatterjee, P., Hoffman, D. L., y Novak, T. P. (2003). Modeling the Clickstream: Implications for Web-Based Advertising Efforts. *Marketing Science*, 22(4), 520–541. Descargado de <http://dx.doi.org/10.1287/mksc.22.4.520.24906> doi: 10.1287/mksc.22.4.520.24906
- Courbet, D., Fourquet-Courbet, M.-P., Kazan, R., y Intartaglia, J. (2014). The Long-Term Effects of E-Advertising: The Influence of Internet Pop-ups Viewed at a Low Level of Attention in Implicit Memory. *Journal of Computer-Mediated Communication*, 19(2), 274–293. Descargado de <http://doi.wiley.com/10.1111/jcc4.12035> doi: 10.1111/jcc4.12035
- Drèze, X., y Hussherr, F. X. (2003). Internet advertising: Is anybody watching? *Journal of Interactive Marketing*, 17(4), 8–23. Descargado de <http://linkinghub.elsevier.com/retrieve/pii/S1094996803701431> doi: 10.1002/dir.10063
- Fruits From Chile. (2014). *Chile and its fruit industry*. Descargado 2015-04-12, de <http://www.fruitsfromchile.com/industry.php>
- Goldstein, D., McAfee, R., y Suri, S. (2013). The cost of annoying ads. *Www*, 459–469.

- Descargado de <http://dl.acm.org/citation.cfm?id=2488429>
- Google. (2016, 27 de 08). *Display benchmarks*. Descargado de <http://www.richmediagallery.com/tools/benchmarks> (Accedido el 27-08-2016)
- Google AdSense. (2015). *Ad placement policies - AdSense Help*. Descargado 2016-01-14, de [#}Ad{ }limit{ }per{ }page](https://support.google.com/adsense/answer/1346295?hl=en)
- Hax, A. (2009). *The delta model: reinventing your business strategy*. Springer Science & Business.
- Hof, R. (2013, 04). *Here's the future of advertising, according to google*. Descargado de <https://www.forbes.com/sites/roberthof/2013/04/10/heres-the-future-of-advertising-according-to-google/> (Accedido el 31-12-2017)
- Hoppe, A., Roxin, A., y Nicolle, C. (2014). Customizing semantic profiling for digital advertising. En R. Meersman y cols. (Eds.), *On the move to meaningful internet systems: Otm 2014 workshops* (pp. 469–478). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Lee, J.-H., Ha, J., Jung, J.-Y., y Lee, S. (2013). Semantic contextual advertising based on the open directory project. *ACM Transactions on the Web*, 7(4), 24:1—24:22. Descargado de <http://www.scopus.com/inward/record.url?eid=2-s2.0-84887481113&partnerID=40&md5=a6a405027183c53168d4a66bf0a4baa4> doi: 10.1145/2529995.2529997
- Mayer, R. (1992). *IDEF0 function modeling*. Air Force Systems Command.
- Montgomery, D. (2012). *Design and analysis of experiments, 8th edition*. John Wiley & Sons, Incorporated. Descargado de <https://books.google.cl/books?id=XQAcAAAAQBAJ>
- Osterwalder, A., y Pigneur, Y. (2010). *Business model generation: a handbook for visionaries, game changers, and challengers*. John Wiley & Sons.
- RStudio Team. (2015). *Rstudio: Integrated development environment for r* [Manual de software informático]. Boston, MA. Descargado de <http://www.rstudio.com/>
- Trademap.com. (2015). *List of exporters for product : 08 edible fruit, nuts, peel of citrus fruit, melons*. Descargado 2015-04-15, de <http://www.trademap.org/tradestat/Country{ }SelProduct.aspx?nvpm=1|||||08|||2|1|1|2|1|1|2|1|1>
- Wu, X., Yan, J., Liu, N., Yan, S., Chen, Y., y Chen, Z. (2009). Probabilistic Latent Semantic User Segmentation for Behavioral Targeted Advertising *. *Proceedings of the Third International Workshop on Data Mining and Audience Intelligence for Advertising*, 10–17. doi: 10.1145/1592748.1592751
- Wu, Z., Xu, G., Zhang, Y., Dolog, P., y Lu, C. (2012). An Improved Contextual Advertising Matching Approach based on Wikipedia Knowledge. *The Computer Journal*, 55(3), 277–292. Descargado de <http://comjnl.oxfordjournals.org/content/55/3/277.abstract> doi: 10.1093/comjnl/bxr056
- Xu, G., Wu, Z., Li, G., y Chen, E. (2014). Improving contextual advertising matching by using Wikipedia thesaurus knowledge. *Knowledge and Information Systems*, 43(3), 599–631. Descargado de <http://link.springer.com/10.1007/s10115-014-0745-z> doi: 10.1007/s10115-014-0745-z
- Yih, W.-t., Goodman, J., y Carvalho, V. R. (2006). Finding advertising keywords on web pages. *Proceedings of the 15th international conference on World Wide Web - WWW '06*, 213. Descargado de <http://dl.acm.org/citation.cfm?id=1135777.1135813> doi: 10.1145/1135777.1135813