



**TESIS DE GRADO EN
INGENIERÍA INDUSTRIAL**

**MARKETING ONLINE:
ESTRATEGIAS INTELIGENTES**

Autor: Sonia Moser

Legajo 46412

Tutor: Ing. Andrés Agres, MSc.

Co-Tutor: Ing. Martín Panelati

2011

RESUMEN

Como usuarios de Internet formamos parte del sistema de la red de búsquedas diariamente con solo abrir un navegador y realizar una consulta. Los enlaces que aparecen sobre el margen derecho son enlaces patrocinados, previamente contratados por un anunciante, que funcionan mediante un mecanismo de subastas.

Cada vez son más las empresas que contratan estos anuncios y cada vez es mayor el precio que están dispuestas a pagar por ellos. El marketing online está ganando importancia en el mercado de la publicidad por su eficiencia en captar potenciales clientes, lograr objetivos concretos y por la gran exposición que tiene.

Este análisis está basado en Google Adwords, el servicio de publicidad de enlaces patrocinados que ofrece Google, sistema de reglas claras y simples pero a la vez complejo por la gran interacción de los actores que lo conforman.

Para lograr esto se explica detalladamente cada uno de los elementos que lo conforman. Luego se definen un grupo de estrategias claras y consistentes y se crea un sistema teórico donde se prueban las mismas.

La más compleja de las estrategias es una optimización global que se ejecuta mediante un modelo de simulación. Esta estrategia obtiene resultados en promedio cuatro veces mayores al resto.

Es aquí donde radica lo más interesante de este análisis: en enfocar un tema que generalmente es analizado desde el punto de vista de la publicidad y del marketing, desde un punto de vista matemático.

DESCRIPTOR BIBLIOGRÁFICO

Esta tesis de grado trata sobre el análisis detallado del sistema de subastas de Google AdWords, herramienta de Google para publicidad mediante enlaces patrocinados.

Para ello primero se revisan los conceptos básicos de la publicidad en internet, modelización, simulación y optimizaciones matemáticas. Luego se expone el contexto desde su comienzo hasta la situación actual sugiriendo también las tendencias futuras. Se explica con detalle el sistema de las subastas y la asignación de precios, y el sistema teórico utilizando hipótesis probadas en publicaciones anteriores. Por último se identifican las estrategias a probar, se explica detalladamente su metodología, se realiza el análisis y se explicitan las conclusiones.

Se obtienen resultados satisfactorios y en algunos casos contra-intuitivos. Además se obtienen conclusiones generales sobre la metodología y las herramientas utilizadas.

Palabras clave: enlaces patrocinados, Google AdWords, subastas, SEM, optimización matemática, simulación.

ABSTRACT

This paper is about Google's online marketing service: Google AdWords. It focuses on understanding how it works and then analyses the result of using different automatic strategies.

It begins with a brief explanation on how Internet Advertisement works and on other concepts used such as modeling, simulation and mathematical optimizations. The context is then exposed for a further understanding on the importance of the theme. Finally the strategies are detailed, tried on many different theoretical scenarios and finally the results are analyzed and conclusions are drawn.

The results are both satisfactory and in some cases counterintuitive. Additionally more general conclusions are obtained regarding the methodology and the tools that were developed.

Keywords: sponsored links, Google AdWords, auctions, SEM, mathematical optimization, simulation.

AGRADECIMIENTOS

Quiero aprovechar estas líneas para agradecer a las siguientes personas por su apoyo y colaboración en este proyecto y en toda esta etapa que con él termina.

A mis tutores Martin Panelati y Andrés Agres y a Juan Pablo Rodriguez Varela por ayudarme en la tesis con su tiempo, dedicación y paciencia.

A mi familia por su gran confianza y apoyo incondicional.

A Sebastián por incentivar me y acompañarme con este proyecto y todos los días.

A mis amigos los mejores compañeros de la vida.

CONTENIDO

1	<u>INTRODUCCIÓN</u>	1
1.1	INTRODUCCIÓN	1
1.2	PROBLEMA Y OBJETIVO	1
1.3	LA IMPORTANCIA DEL TEMA	2
1.4	ESQUEMA DEL PROYECTO	2
2	<u>MARCO TEÓRICO</u>	3
2.1	¿CÓMO FUNCIONA LA PUBLICIDAD EN INTERNET?	3
2.1.1	MOTORES DE BÚSQUEDA DE INTERNET: INTRODUCCIÓN A LA CADENA DE VALOR	3
2.1.2	TIPOS DE PUBLICIDAD EN INTERNET	4
2.1.3	SISTEMAS DE PAGO PARA LA PUBLICIDAD ONLINE	8
2.1.4	SISTEMAS DE PRECIO PARA EL SISTEMA DE CPC (COSTO POR <i>CLICK</i>)	8
2.2	PUBLICIDAD EN INTERNET: FOCO EN GOOGLE	9
2.2.1	NIVEL DE CALIDAD (QUALITY SCORE)	10
2.2.2	CÓMO AFECTA EL NIVEL DE CALIDAD	11
2.2.3	ESTRUCTURA DE LAS CUENTAS DE GOOGLE ADWORDS	12
2.3	OPTIMIZACIÓN MATEMÁTICA	13
2.3.1	CLASIFICACIÓN DE MÉTODOS DE OPTIMIZACIÓN	13
2.3.2	MÉTODOS META-HEURÍSTICOS DE OPTIMIZACIÓN	15
2.3.3	OPTIMIZACIONES BASADAS EN COLONIAS DE HORMIGAS (ACO: ANT COLONY OPTIMIZATIONS)	16
2.4	SIMULACIÓN Y MODELIZACIÓN	17
2.4.1	EL PROCESO DE MODELIZACIÓN	18
2.4.2	SIMULACIÓN POR AGENTES	19
3	<u>CONTEXTO</u>	20
4	<u>METODOLOGÍA</u>	22
4.1	ESPECIFICACIÓN DEL SISTEMA TEÓRICO	22
	INPUTS DEL SISTEMA	22
	OUTPUTS DEL SISTEMA	23
4.1.1	COSTO	24
4.1.1	INGRESO	26
4.1.2	VOLUMEN	27
4.1.3	PRESUPUESTO	28
4.1.4	CÁLCULOS	28
4.1.5	AJUSTE DEL RESULTADO AL PRESUPUESTO	29
4.1.6	ÍNDICE DE CALIDAD	29
4.1.7	HORIZONTE TEMPORAL DE PLANEAMIENTO	30
4.2	IDENTIFICACIÓN DE LAS ESTRATEGIAS	31

TIPO 1: ESTRATEGIAS EQUITATIVAS	31
TIPO 2: ESTRATEGIAS DE MAXIMIZACIÓN POR PALABRA	33
TIPO 3: ESTRATEGIAS DE RANKING	35
TIPO 4: ESTRATEGIAS DE OPTIMIZACIÓN GLOBAL	37
<u>5 METODOLOGÍA APLICADA & ANÁLISIS</u>	50
5.1 ANÁLISIS GENERAL DE LAS SALIDAS	51
5.2 SEGMENTACIÓN DE LAS CORRIDAS	53
5.3 ANÁLISIS POR TIPO DE ESTRATEGIA	54
5.3.1 ESTRATEGIA TIPO 1: ESTRATEGIAS CUANTITATIVAS.....	54
5.3.2 ESTRATEGIA TIPO 2: ESTRATEGIAS DE MAXIMIZACIÓN POR PALABRA	55
5.3.3 ESTRATEGIA TIPO 3: ESTRATEGIAS DE RANKING.....	57
5.3.4 ESTRATEGIA TIPO 4: OPTIMIZACIÓN GLOBAL	60
5.4 ANÁLISIS MÁS ALLÁ DEL RESULTADO	61
5.4.1 TIEMPO.....	61
5.4.2 COMPLEJIDAD	61
5.5 RESUMEN DEL ANÁLISIS	62
<u>6 CONCLUSIONES Y FUTURAS LÍNEAS DE INVESTIGACIÓN</u>	63
6.1 CONCLUSIONES DE LA TESIS.....	63
6.2 CONCLUSIONES GENERALES DE LA METODOLOGÍA UTILIZADA	64
6.3 LIMITACIONES DEL ESTUDIO.....	65
6.4 FUTURAS LÍNEAS DE INVESTIGACIÓN	66
<u>7 BIBLIOGRAFÍA</u>	68
<u>8 ANEXO</u>	70
8.1 RESULTADOS COMPLETOS.....	70

1 INTRODUCCIÓN

En este Capítulo se analizará el objetivo del Proyecto.

1.1 Introducción

Cada vez más empresas están incursionando en prácticas de marketing en Internet, destinándoles un mayor presupuesto con el objetivo de captar a un mayor número de potenciales clientes por este medio. Las tendencias muestran que la publicidad en internet viene creciendo sostenidamente en todo el mundo.

Invertir en publicidad online crea una ventaja competitiva inmediata. No hacerlo implica un gran resigo. Y si bien las herramientas ofrecidas por los buscadores son muy fáciles de utilizar y no requieren de intermediarios, puede resultar difícil lograr los objetivos propuestos.

1.2 Problema y objetivo

Las herramientas ofrecidas por los buscadores, al igual que muchos de los servicios ofrecidos en Internet, son extremadamente simples de utilizar pero funcionan a través de un mecanismo absolutamente complejo.

Esto se debe a que internet es una red abierta completamente dinámica. Cualquiera con acceso a internet puede realizar una búsqueda y al hacerlo se convierte en un actor que modifica al sistema sin necesariamente comprender su funcionamiento. Esto hace que sea difícil entender y predecir el comportamiento de los usuarios.

Utilizar publicidad por internet es simple. Pero lograr objetivos determinados como por ejemplo captar un público específico, gastar el presupuesto destinado para una campaña, lograr el retorno esperado entre tantos otros puede ser muy complicado. Más aún si lo que se busca es utilizar el presupuesto de manera eficiente.

De ahí radica **el objetivo de este proyecto que consta en identificar y comparar diferentes estrategias de inversión en una de las principales ramas del marketing online: la publicidad mediante enlaces patrocinados en la red de búsqueda.** En concreto se acota el análisis a Google Adwords, el servicio que ofrece Google, el buscador que tiene más del 85% del mercado de búsquedas.

Esta herramienta funciona a través de un método de subastas. El resultado de cualquier subasta depende del comportamiento de todos los ofertantes en conjunto. Si bien hay reglas absolutas, las cantidades ofertadas adquieren una naturaleza relativa. Una oferta será la máxima en función a lo que subasten los demás.

Por todas estas razones se concluye que este es un problema complejo que depende de muchos actores que interactúan en un sistema completamente dinámico.

1.3 La importancia del tema

En sólo diez años, del año 2000 al año 2010, los usuarios de internet han crecido un 450% en todo el mundo¹. Los buscadores de Internet generan millones de momentos relevantes diarios al conectar a consumidores que necesitan productos y servicios específicos, con empresas y organizaciones que ofrecen precisamente esos productos y servicios.

Mientras diferentes medios de comunicación como la televisión, los periódicos y la radio vienen perdiendo cuota del mercado, las tendencias muestran que la publicidad en Internet viene creciendo sostenidamente en todo el mundo. En 2010 la inversión publicitaria en Internet en Estados Unidos superó por primera vez la inversión en los periódicos creciendo un 13,9% con respecto al año anterior². No sólo está creciendo el número de ofertantes sino que también el presupuesto que estos manejan. Sólo en junio de 2010, 47 empresas de Estados Unidos gastaron más de de 1 millón de dólares cada una³ en publicidad en Internet.

Todos los pronósticos indican que la publicidad online seguirá creciendo y ganando cuota en el mercado publicitario. Por un lado por el crecimiento de la cantidad de usuarios de internet en el mundo, por el otro porque según la BBC del Reino Unido cada dólar gastado en marketing online obtiene un retorno sobre la inversión mayor que cada dólar invertido en medios tradicionales como televisión, periódicos y radio⁴.

1.4 Esquema del Proyecto

Este estudio busca comprender diferentes estrategias para realizar subastas en Google AdWords de forma óptima en pos de conclusiones útiles y extrapolables del sistema teórico al mundo real. El esquema a seguir es el siguiente:

- i. Identificación y explicación detallada de las estrategias.
- ii. Codificación de las estrategias mediante reglas de decisión claras y automáticas.
- iii. Creación de un sistema teórico.
- iv. Aplicación de las estrategias en el sistema teórico.
- v. Análisis y conclusiones.

¹Fuente: Internet World Stats: <http://www.internetworldstats.com/stats.htm>, junio 2011.

² Fuente: www.elperiodico.com, junio 2011.

³ Fuente: un documento filtrado de Google: <http://www.readwriteweb.es/economia/inversion-publicitaria-adwords-grandes-marcas/>

⁴Fuente: http://www.bbc.co.uk/blogs/technology/2008/04/google_how_big_is_too_big.html, abril 2008

2 MARCO TEÓRICO

Se proveerá la teoría detrás de los conceptos más importantes que se usarán en el desarrollo del proyecto explicando de manera detallada el funcionamiento de la publicidad en Internet y otros conceptos como simulación por agentes y optimizaciones matemáticas.

2.1 ¿Cómo funciona la publicidad en internet?

2.1.1 Motores de búsqueda de Internet: introducción a la cadena de valor.

Más del 30% de la población mundial tiene acceso a Internet. Esto implica más de 2 billones de usuarios en todo el mundo navegando por la gran *World Wide Web*.

Pero, ¿cuál es el negocio de los motores de búsqueda? ¿En qué parte de este gran intercambio de información es que el motor de búsqueda genera valor y gana dinero? ¿De qué le sirve a Google recibir cientos de millones de búsquedas diarias si al devolverle a sus usuarios los resultados de las búsquedas no recibe dinero?

La respuesta es publicidad. El principal agregado de valor de los motores de búsqueda es la publicidad en internet. Si bien los buscadores atraen a sus usuarios por la calidad, rapidez y relevancia de los resultados de sus búsquedas, generan dinero por medio de la publicidad.

Se llama SEM (*search engine marketing*) al conjunto de herramientas que ofrecen los propios buscadores para publicitar a los sitios en sus redes permitiendo que éstos manejen su marketing online sin necesidad de intermediarios.

Como en cualquier otro medio de publicidad, cuanto mayor sea el caudal de personas en contacto con el medio, más atractivo será para quien publique su anuncio. De modo que a mayor cantidad de usuarios más atractivo será el buscador para los anunciantes, mayor será la demanda de anunciantes, mayor será el

dinero invertido por los anunciantes y mayor serán los ingresos del buscador. Como consecuencia mayor será el presupuesto que el buscador podrá reinvertir en mejoras y de hacerlo, más atractivo será el buscador para sus usuarios.

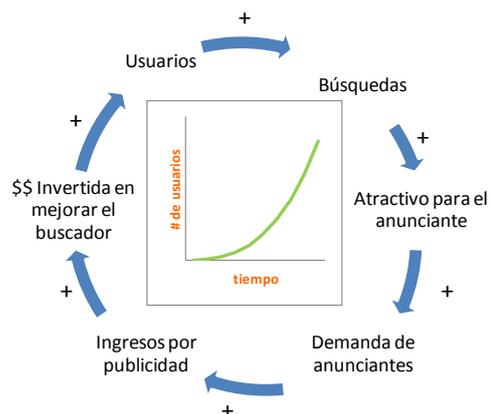


Diagrama 1: Diagrama causal simplificado: publicidad en internet

Este diagrama causal es una gran simplificación de la dinámica de adopción de usuarios. Pero de él se desprenden dos conclusiones importantes:

1. A mayor caudal de usuarios, más atractivo será para los anunciantes y por ende más ingresos por publicidad.
2. El diagrama muestra una relación reforzadora. El reforzador puede actuar en forma positiva, como lo muestra la forma de la curva de cantidad de usuarios, pero también puede actuar en el sentido contrario. Hay un gran riesgo de que la misma publicidad actúe en forma negativa en la imagen del buscador, haciendo que disminuya la satisfacción del usuario y generando así lo contrario a lo deseado. Si disminuye la cantidad de usuarios el buscador pierde atractivo para el anunciante y disminuye la inversión en publicidad.

2.1.2 Tipos de publicidad en Internet

La publicidad en internet consta de dos elementos principales: un **anuncio** y una **página de destino** (*landing page*). Cuando el usuario realiza un *click* sobre el anuncio, será enviado a la página de destino.

Hay diferentes maneras de publicitar un anuncio, a continuación aparecen algunas de ellas. Pero es importante entender que el objetivo principal de dichos anuncios es aumentar el tráfico de usuarios en las páginas de destino. Luego si se logra o no vender el producto o servicio, crear la imagen de marca deseada, o cualquier otro objetivo que se pueda tener, depende exclusivamente de las características de la página. Por ejemplo, si se busca vender un producto por internet una vez que el usuario llega a la página de destino, dependerá de que ésta sea atractiva, fácil de utilizar, veloz, con un sistema de pago que genere confianza y sea simple, entre tantas otras características.

Queda claro entonces que ambos elementos de la campaña publicitaria deben ser tenidos en cuenta. Sin un anuncio atractivo no se logrará tener el tráfico deseado en la página de destino. Pero, si la página de destino no tiene las características necesarias, invertir en publicidad no tiene sentido alguno.

Algunas de las formas de colocar anuncios:

Publicidad con enlaces patrocinados: un enlace patrocinado es aquel que ha sido previamente contratado por el anunciante. Existen dos situaciones en las que un usuario se encontrará con enlaces patrocinados navegando por internet. La primera es mientras realiza una búsqueda (red de búsquedas), la segunda es mientras navega por las diferentes páginas web (red de display).

- **Enlaces patrocinados en la red de búsquedas:** Cada vez que un usuario realiza una búsqueda en internet, obtiene una lista de enlaces de búsqueda

(también llamados enlaces orgánicos) y una lista de enlaces patrocinados asociados a la búsqueda realizada. En el siguiente ejemplo aparecen los elementos principales de una búsqueda.



- **Un buscador:** Google
- **Un usuario:** quien realiza la búsqueda
- **Una consulta:** “Buenos Aires outdoors”

Y la búsqueda dará como resultado



Ilustración 1: Resultado de una búsqueda

- **Enlaces orgánicos** listados en orden de relevancia según el buscador con el objetivo de satisfacer al usuario con su búsqueda, es decir intentando posicionar los enlaces más significativos en primer lugar. Algunas variables que influirán en este posicionamiento son el título del sitio, su contenido, la relevancia del título en el resto del sitio y la cantidad de enlaces a sitios externos.
- **Enlaces patrocinados** listados según la consulta que se ha realizado, cuánto está dispuesto a pagar el anunciante por cada *click* y la calidad del anuncio (explicado en la sección 2.2).
- **Enlaces patrocinados en la red de display:** Cada vez es mayor el número de páginas web que incluyen enlaces patrocinados para generar

ingresos. Se llama red de display al conjunto de estos sitios. Hay dos modalidades en las que una página web consigue y publica enlaces patrocinados.

- i. Hay páginas web que tienen sus propios sistemas de publicidad. En este caso el contacto es directo entre la página web y el anunciante. Para que esto suceda la página web debe tener una determinada estructura que permita el



Ilustración 2: Ejemplo de enlaces patrocinados en Facebook

desarrollo del sistema y para que esto sea posible debe tener un gran flujo de visitas. Entre estas páginas se encuentran redes sociales, periódicos y páginas que ofrecen servicios. Algunos ejemplos son Facebook, Youtube, Myspace, La Nación, Cuevana, etc. La ilustración 1 muestra un ejemplo de este tipo de publicidad en Facebook.

- ii. La gran mayoría de páginas web no tienen sus propios sistemas de publicidad. Algunos buscadores ofrecen sistemas para estos casos: la página web interesada debe registrarse y el anunciante debe permitir que sus enlaces patrocinados aparezcan en la red de display. Luego parte del ingreso de la publicidad para la página que la publica y la otra parte va para el buscador. Por ejemplo Google ofrece este sistema mediante *Google AdSense*. A continuación se ve el ejemplo de un blog llamado “El blog Salmón” que contiene enlaces patrocinados de Google.



Ilustración 3: Ejemplo de enlaces patrocinados en Blogs

La gran ventaja de publicar enlaces patrocinados en la red de display es que los anuncios pueden segmentarse. En las redes sociales los usuarios tiene

perfiles que generalmente contienen mucha información que va desde edad y sexo hasta gustos y actividades cotidianas. En las páginas especializadas como los blogs, las temáticas suelen ser muy acotadas de modo que se puede conocer el perfil del usuario promedio.

Esta es información muy valiosa para el anunciante ya que le permite apuntar sus campañas específicamente al público que le interesa permitiéndole utilizar su dinero de manera más eficiente.

Banners: son imágenes que pueden ser dinámicas o estáticas que se colocan en un lugar sobresaliente de una página web. Normalmente el precio del banner es proporcional al tráfico de la página web donde se publica

Ventanas emergentes (publicidad *pop-up* o *pop-under*): son ventanas que se abren repentinamente frente al usuario. Como no requieren de un permiso del usuario para hacerlo pueden ser molestas y criticadas. Las ventanas emergentes *pop-ups* son las que aparecen frente a la ventana activa mientras que las *pop-under*s aparecen debajo de la ventana activa y recién son vistas cuando el usuario cierra la ventana.

Publicidad en anuncios clasificados gratis en sitios de internet: hay sitios de internet que se dedican a publicar anuncios clasificados de forma gratuita. Como tienen un gran contenido de clasificados los motores de búsqueda suelen darles mucha importancia y logran salir en las primeras posiciones de las búsquedas. Por esta razón son sitios que reciben muchas visitas y por ende son atractivos para terceros para exhibir sus publicidades. Entonces si bien la publicidad de los clasificados es gratis, estos sitios logran generar ingresos a través de la publicidad de terceros.

Publicidad por correo electrónico: el objetivo de esta forma de publicidad es el mismo que el de la publicidad por correo, la empresa elige a los clientes o potenciales clientes con los cuales desea establecer una relación de comunicación y lo hace por medio de correos electrónicos. Pero fue tal el abuso por parte de ciertas empresas, que en la actualidad muchos sistemas de correo electrónico tienen filtros *anti-spam* que impiden la llegada de este tipo de correo electrónico. Todavía es posible lograr campañas efectivas por este medio, pero para hacerlo el anunciante debe lograr el equilibrio justo para no saturar a los receptores.

Publicidad en videos de Internet: hay diferentes sitios en internet que permiten a sus usuarios subir videos, compartirlos y así difundirlos. Algunos de ellos muestran anuncios durante su transmisión. El gran potencial de este tipo de publicidad radica en que cuando un video captura la atención de un usuario logra mantenerlo concentrado por un tiempo largo. Esto tiene un enorme valor para el anunciante ya que es más probable que el usuario perciba el anuncio que en muchos de los otros ejemplos. De la misma manera esto tiene un gran riesgo para el sitio ya que los anuncios pueden ser molestos para los usuarios.

Publicidad gratis en Google: el ejemplo de este tipo de publicidad es un servicio que ofrece Google, que se llama *Google Local*. Las empresas deben registrarse en *local business center de Google* y cuando un usuario realiza una búsqueda e indica una localidad aparece en el resultado un mapa de *Google maps* marcando la empresa en cuestión. Este servicio de Google está presente sólo en algunos países, y sirve únicamente para publicitar empresas y servicios que tienen un punto físico a la calle. Pero tiene una gran ventaja y es que es gratis y muy efectivo ya que sitúa los resultados de este tipo de búsquedas en los primeros lugares.

2.1.3 Sistemas de pago para la publicidad online

Existen diferentes modelos de pago para la publicidad online. Los principales son:

CPC (costo por click): también conocido como PPC (*pay per click*) este es el sistema más común en la actualidad. La mayoría de los sistemas de publicidad en internet han tendido a este sistema de pago. Tal como el nombre lo implica, en este sistema el anunciante sólo paga si alguien realiza un *click* sobre el anuncio. Cuánto pagará dependerá del sistema de precio pero lo importante es que si un anuncio es impreso pero no recibe *clicks* el costo para el anunciante es nulo.

CPM (costo por mil): también conocido como CPT (*cost per thousand*) en este sistema se paga por impresiones. Cada vez que un anuncio aparece en la pantalla de un usuario se genera una impresión pero lógicamente esto no implica que el usuario vea el anuncio. Generalmente se cobra un monto fijo cada mil impresiones.

CPV (costo por visitante): se paga cuando un determinado tipo de usuario ve el anuncio. En la práctica, se paga por impresión, pero los anuncios se imprimen únicamente en determinadas búsquedas o páginas webs intentando así segmentar al usuario. Se utiliza principalmente con *pop-ups* y *pop-unders*.

CPA (costo por adquisición): el anunciante paga únicamente si se efectúa la venta (u otro tipo de objetivo como por ejemplo votación, reserva, etc.) de aquello que anunciaba. Quien publica el anuncio corre todo el riesgo, y asume que no habrán ineficiencias en la página web del anunciante.

2.1.4 Sistemas de precio para el sistema de CPC (costo por click)

Una vez fijado el sistema de pago, debe definirse el sistema de precio. Es decir, ¿cuánto pagará el anunciante cada vez que deba pagar? Normalmente se utiliza uno de los dos sistemas explicados a continuación:

Hay sistemas de publicidad que tienen **montos fijos:** se paga un precio que por lo general es proporcional al flujo de visitantes de la página. Una página que utiliza este sistema es www.travelocity.com.ar y Facebook también presenta la opción de cobrar montos fijos por impresiones.

El otro sistema, que en la actualidad es el más utilizado para los enlaces patrocinados, es la determinación de precios mediante sistemas de **subastas**. De modo que el precio que pagará el anunciante es relativo a los que el resto de los interesados oferte. El sistema más común es la **subasta de sobre cerrado**: cada anunciante define cuánto está dispuesto a pagar (por *click*, por impresión, por acción, etc). Quien oferte más aparecerá en primer lugar, quien oferte la segunda posición aparecerá segundo y así sucesivamente hasta que se llenen todos los cupos de enlaces patrocinados. Además la subasta puede ser **de primer precio o de segundo precio**. En el primer caso cada anunciante pagará su propia oferta, en el segundo caso el primero pagará lo que ofertó el segundo, y así progresivamente. Lo más común en los enlaces patrocinados son las **subastas de sobre cerrado de segundo precio**.

2.2 Publicidad en Internet: Foco en Google

Google comenzó a publicar anuncios publicitarios en enero del año 2000 utilizando anuncios de textos. En ese momento ya eran muy populares los banners y otras formas de anuncios obstructivos (como *pop-ups* y *pop unders*). Google cobraba los anuncios de texto con un sistema CPM (costo por miles) y con un precio fijo.

En octubre del 2000 arrancó Google AdWords todavía con un sistema de CPM. En 2001 el beneficio de Google por publicidad había alcanzado los 70 millones de dólares. Sin embargo en ese momento Overture⁵ había ganado 288 millones de dólares con su sistema de publicidad que utilizaba el sistema de CPC y fijaba los precios mediante subastas.

En febrero de 2002 Google lanzó su nueva versión de Adwords que había incorporado los conceptos utilizados por Overture: utilizaba un sistema de CPC y de subastas. El modelo de negocio utilizado era el que muestra el Diagrama 2.



Diagrama 2: modelo de negocio de Google Adwords

Pero Google no copió a Overture enteramente; agregó además una innovación gracias a la cual en el 2010 tuvo ganancias de más de 28.000 millones de dólares.

La gran innovación de Google fue introducir una nueva variable en la ecuación del posicionamiento: **relevancia**. El algoritmo de ranking comenzó a contemplar el indicador de CTR (*click through rate*) como medida de relevancia del anuncio. El CTR es un indicador que calcula de todas las veces que el anuncio es impreso,

⁵OvertureServices: desde 2001 es el nombre del sitio GoTo.com que luego fue adquirido por Yahoo y pasó a llamarse *YahooSearch Marketing*)

cuántas veces recibe un *click*. Así, si el CTR es bajo se puede concluir que la relevancia de ese anuncio es baja.

El modelo de Overture sólo contemplaba la subasta para posicionar sus anuncios de modo que un anunciante podía comprar su lugar en las primeras posiciones: únicamente debía pagar el precio correspondiente sin importar si el anuncio era relevante para esa búsqueda. El gran problema de este modelo es que si un anuncio es irrelevante no recibirá *clicks* por más que se encuentre en la primera posición de todas las búsquedas. Si el anuncio no recibe *clicks* el buscador (en este caso Overture) no genera ingresos.

2.2.1 Nivel de calidad (quality score)

Con el tiempo Google continuó desarrollando su algoritmo de ranking. Hoy no sólo mira la relevancia del anuncio sino que existe un **nivel de calidad** para cada anuncio cuyos factores principales son:

- **CTR (click through rate)** : porcentaje de *clicks* que recibe el anuncio sobre el total de impresiones (de todas las veces que aparece el anuncio, cuantas veces recibe un *click*). Google siempre utiliza los comentarios y las experiencias de los usuarios con el objetivos de maximizar su conformidad, y este indicador es un reflejo de la trascendencia que tiene el anuncio para los usuarios.
- **Relevancia**: mira la relación entre la palabra clave con y los anuncios del grupo de anuncios. También mira la relación entre de la palabra clave y del anuncio correspondiente con respecto a la consulta de búsqueda. Por medio de este indicador Google busca asegurarse que los usuarios obtengan anuncios útiles con sus búsquedas. Además Google evita que anunciantes logren publicar anuncios mediante palabras de búsqueda que no tengan relevancia alguna con el producto o servicio que venden.
- **Calidad de la página de destino**: un anuncio sólo será útil para un usuario si su página de destino lo ayuda a encontrar la información que éste necesita. La misma debe ser:
 - Relevante
 - Con contenido original
 - Fácil de navegar, con tiempos de carga cortos y con pocos *pop-ups* y *pop-unders*.
 - La naturaleza del servicio o producto ofrecido debe ser transparente y claro.
- **Otros factores** menos importantes son:
 - El CTR histórico de la cuenta (se mide a través de todos los anuncios y todas las palabras de la cuenta)

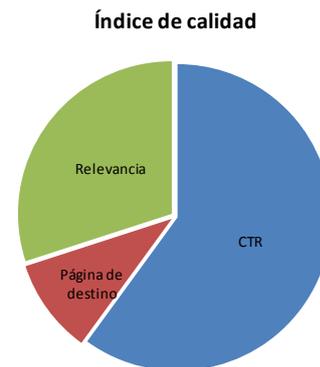


Gráfico 1: Componentes principales del Índice de Calidad

- El rendimiento histórico de la cuenta en la región geográfica en la que aparece el anuncio.

2.2.2 Cómo afecta el nivel de calidad

Según Hall Varian, economista de Google, para que una subasta sea exitosa deben reconciliarse los intereses de las tres partes involucradas.

El anunciante desea mostrar anuncios relevantes, para que los usuarios decidan realizar *clicks* sobre ellos.

Los usuarios desean ver anuncios relevantes, anuncios irrelevantes son considerados molestos.

Google quiere crear una buena experiencia, tanto para los anunciantes como para los usuarios para que vuelvan a utilizar el servicio en el futuro.

De modo que el índice de calidad incide sobre los tres actores: ayuda a aumentar los ingresos de Google haciendo que los anuncios irrelevantes no puedan estar en las posiciones más altas, cuida la experiencia del usuario al vigilar la relevancia de los anuncios con los que se topará y cuida la experiencia de los anunciantes privilegiando a los anuncios más relevantes.

Google AdWords funciona mediante una **subasta de sobre cerrado de segundo precio**. El ejemplo a continuación explica en detalle el funcionamiento de Google AdWords en versión simplificada y luego con el efecto del índice de calidad.

Cómo funciona Google AdWords: Ejemplo simplificado

En este ejemplo hay tres anunciantes que subastan por una de las tres posiciones posibles de una misma palabra clave.

Anunciante	Precio máximo (\$)	Posición	Precio real (\$)
Pedro	6	1 ^{ra}	5
Tomás	5	2 ^{da}	1
Agustín	1	3 ^{ra}	Precio mínimo

Tabla 1: Ejemplo simplificado

Como se trata de una subasta de segundo precio Pedro, quien está dispuesto a pagar 6 se colocará en la primera posición pero pagará 5, el precio que subastó Tomás.

Cómo funciona Google AdWords: Ejemplo completo

Ahora al ejemplo anterior se le agrego el índice de calidad, que va de 1 hasta 10.

Donde Ad Rank = Precio máximo x nivel de calidad

Anunciante	Precio máximo (\$)	Nivel de Calidad	Ad rank	Posición
Pedro	6	2	12	2 ^{da}
Tomás	5	4	20	1 ^{ra}
Agustín	1	6	6	3 ^{ra}

Tabla 2: Ejemplo completo

Se ve que con las misma subastas del ejemplo anterior, cambiaron las posiciones debido al nivel de calidad. Ahora, **¿cómo se define el precio real que pagará el anunciante por *click*?**

Se mantiene el mismo principio del caso anterior: se paga el precio mínimo necesario para mantener la posición.

Entonces el producto del precio que pagará Pablo y su índice de calidad debe igualar al producto de la oferta máxima de Tomás y su índice de calidad.

$$Precio_{Tomás} \times Índice\ Calidad_{Tomás} = PrecioMáximo_{Pedro} \times Índice\ Calidad_{Pedro}$$

$$Precio_{Tomás} = \frac{Precio\ Máximo_{Pedro} \times Índice\ Calidad_{Pedro}}{Índice\ Calidad_{Tomás}} = \frac{Ad\ Rank_{Pedro}}{Índice\ Calidad_{Tomás}}$$

$$Precio_{Tomás} = \frac{Ad\ Rank_{Pedro}}{Índice\ Calidad_{Tomás}} = \frac{12}{4} = 3$$

$$Precio_{Pedro} = \frac{Ad\ Rank_{Agustín}}{Índice\ Calidad_{Pedro}} = \frac{6}{2} = 3$$

$$Precio_{Agustín} = Precio\ mínimo$$

2.2.3 Estructura de las cuentas de Google Adwords

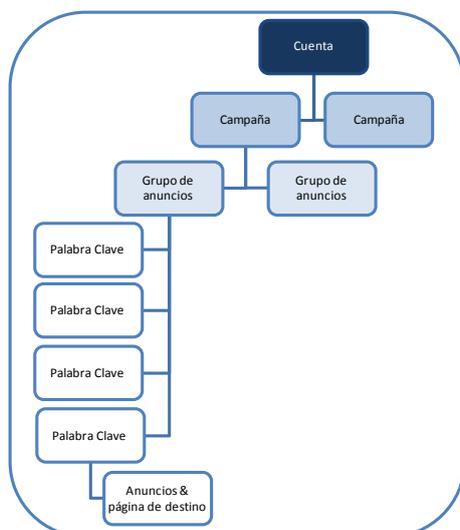


Diagrama 3: estructura de una cuenta de Google Adwords

La estructura de una cuenta de Google Adwords es la siguiente:

Cada campaña tiene asignado un presupuesto diario. A medida que los anuncios reciben *clicks* el presupuesto se va consumiendo. En el momento en que se gasta todo el presupuesto los anuncios dejan de aparecer.

No es posible asignar un presupuesto fijo a cada grupo de anuncios, tampoco puede definirse un presupuesto por palabra clave; siempre se define el presupuesto de toda la campaña.

Existen diferentes maneras de organizar un grupo

de anuncios. Se puede tener un anuncio asociado a una palabra clave, o se puede tener un grupo de anuncios asociado a una palabra clave o a un grupo de palabras claves. Además cada anuncio está asociado a una página de destino que puede repetirse o no. Más anuncios y páginas de destino implican una mayor inversión pero permiten ganar un mayor grado de detalle.

2.3 Optimización matemática

En matemática una optimización tiene como objetivo encontrar la mejor solución para un problema entre un conjunto de soluciones posibles. Los problemas de optimización están compuestos por los siguientes elementos:

- Una **función objetivo**: es la expresión matemática del sistema que se desea maximizar o minimizar para así lograr la solución óptima. Algunos ejemplos de funciones objetivo son: minimización de costos variables de un sistema de transporte urbano, maximización de ingresos por venta de una planta de producción farmacéutica, minimización de riesgo en un portafolio de inversiones, minimización de distancias en la red de distribución.
- **Variables de decisión**: son aquellas que afectan el valor de la función objetivo en forma directa. Sus valores están bajo nuestro control e influyen en el desempeño del sistema. La solución de un problema de optimización implica la determinación de los valores óptimos de cada una de estas variables.
- **Restricciones**: son el conjunto de reglas que ciertas variables están obligadas a cumplir. Están expresadas en forma de ecuaciones e inecuaciones. Algunos ejemplos son las potencias máximas y mínimas de un grupo de generación, la cantidad de horas semanales que puede trabajar un hombre, la cantidad de camiones disponibles por hora en una flota de distribución, etc.

En concreto para resolver un problema de optimización se crea un modelo matemático que caracteriza todas las relaciones que existen en el sistema real, se define la función objetivo a optimizar y procurando respetar las restricciones se buscan los valores de las variables de decisión que maximicen o minimicen la función objetivo.

Los métodos de optimización pueden ser clasificados de diferentes maneras.

2.3.1 Clasificación de métodos de optimización

Existen muchos tipos y métodos de optimizaciones. A continuación se explican algunos de ellos.

2.3.1.1 Según las características de la función objetivo y las restricciones.

Optimización lineal: cuando tanto la función objetivo como las restricciones son lineales. Se resuelve con algoritmos simples de álgebra lineal, los más comunes son el método simplex y el método dual.

Optimización lineal con variables enteras: es un caso específico del caso anterior, en el que algunas de las variables de decisión deben tomar valores enteros.

- Optimización lineal entera pura: cuando todas las variables de decisión deben tomar valores enteros.
- Optimización lineal entera binaria: cuando las variables sólo pueden tomar valores de 1 y 0.
- Optimización lineal entera mixta: cuando algunas variables son enteras, de estas quizás algunas pueden ser binarias y el resto son continuas.

Para resolver estas optimizaciones normalmente se parte de la solución de una optimización lineal pero no sirve redondear estos resultados; hay que aplicar algoritmos para hallar los óptimos enteros. El método más utilizado se llama *Branch and bound*, un método que parte de una optimización lineal, e iterativamente agrupa las posibles soluciones y descarta los grupos no óptimos.

Optimización no lineal: cuando la función objetivo y/o las restricciones contienen relaciones no-lineales.

Optimización estocástica: cuando algunas de las restricciones o parámetros dependen de variables aleatorias.

2.3.1.2 Según el tipo de solución

Existen dos grande grupos de solución que dependen tanto de la naturaleza del problema y su complejidad, como de la cantidad de recursos que se le asignarán a encontrar su solución.

Optimizaciones completas: los algoritmos completos garantizan encontrar una solución óptima en un tiempo limitado. Se aplican a problemas que buscan encontrar la mejor solución dentro de un área finita de búsqueda y utilizan métodos que explotan todas las soluciones. La gran ventaja es la garantía de que la solución encontrada será un óptimo global. Sin embargo el tiempo computacional que requieren estos algoritmos suele crecer en forma exponencial a medida que se agregan variables al problema de modo que los recursos necesarios para obtener optimizaciones completas pueden resultar demasiado elevados para el propósito inicial.

Optimizaciones aproximadas: estos algoritmos encuentran buenas soluciones en un tiempo razonables pero no garantizan soluciones óptimas. Son métodos que exploran el espacio de búsqueda de manera eficiente pero corren el riesgo de estancarse en máximos locales en lugar de encontrar el máximo global. En las últimas dos décadas ha emergido un nuevo método aproximado llamado **meta-**

heurística que combina métodos heurísticos⁶ básicos en sistemas de mayor complejidad logrando alcanzar mayor eficiencia y efectividad en la manera en que recorre el espacio de búsquedas.

2.3.2 Métodos Meta-heurísticos de optimización

Son métodos iterativos que combinan de forma inteligente una heurística básica y la aplican en un sistema más complejo. Normalmente estos métodos imitan fenómenos y comportamientos observados en la naturaleza.

Estos algoritmos tienen gran éxito en la práctica, y son fáciles de utilizar y adaptar. Sin embargo son aproximados y son probabilísticos (es decir que no son deterministas). Es por esto que generalmente se aplican cuando otros métodos no son posibles de implementar, cuando no existen heurísticas básicas que brinden soluciones satisfactorias, o cuando no se necesita una solución exacta.

Para que un algoritmo de búsqueda sea efectivo, debe encontrar un balance entre la **explotación** del espacio y la **exploración** del mismo. Estos son dos objetivos contrapuestos ya que la explotación se logra intensificando el esfuerzo empleado en la búsqueda en la región actual mientras que la exploración se logra diversificando el esfuerzo en diferentes regiones del espacio. Cuánto esfuerzo se le dedica a cada característica afectará no sólo el tiempo que tomará la búsqueda sino que también la solución a la que se llegará.

Algunos ejemplos de aplicación son optimización combinatoria, minería de datos, optimización en ingeniería, bio-informática, modelado de sistemas, etc.

Algunas meta-heurísticas parten de búsquedas locales y agregan algoritmos para evitar quedar atrapadas en óptimos locales. Pueden ser con **heurísticas constructivas**, que parten de una solución inicial vacía y va agregando componentes hasta llegar a la solución final; o **heurísticas basadas en trayectorias**, que parten de una solución inicial y en forma iterativa trata de reemplazarla por una solución mejor. Algunos ejemplos de estos métodos son búsqueda tabú, GRASP, y enfriamiento simulado.

Otras meta-heurísticas son **métodos basadas en poblaciones** que por lo general utilizan comportamientos observados en la naturaleza como base de sus algoritmos. Se identifican áreas de alto potencial realizando muestreo del espacio de soluciones y generando nuevas soluciones a partir de combinaciones de las anteriores. Algunos ejemplos son algoritmos genéticos y optimizaciones basadas

⁶La heurística es la capacidad que tiene un sistema para realizar innovaciones positivas sobre sí mismo. Según la Real Academia Española: *heurístico/ca: en algunas ciencias, manera de buscar la solución de un problema mediante métodos no rigurosos, como por tanteo, reglas empíricas, etc.* En concreto, los métodos heurísticos son aquellos que convergen a una solución de manera sistémica.

en colonias de hormigas (ACO: ant colony optimization) u otras optimizaciones de inteligencia de enjambre (*swarm intelligence*).

2.3.3 Optimizaciones basadas en colonias de hormigas (ACO: ant colony optimizations)

Este método es muy útil para resolver problemas de optimización combinatoria. Utiliza un algoritmo está basado en la metodología que utilizan las hormigas en la realidad para encontrar los caminos entre fuentes de comida y sus nidos.

Cuando una hormiga sale de su nido en busca de comida deposita una substancia denominada feromona. La feromona se evapora con el tiempo de modo que la concentración de feromona en un camino depende tanto de la cantidad de hormigas que pasaron por él como del tiempo que transcurrió desde que fue depositada. Caminos con mayor concentración de feromona tienen una mayor probabilidad de ser elegidos por otras hormigas que caminos con menos feromona.

Al principio las hormigas que no encuentran caminos directos suelen morir. Pero a medida que más hormigas salen a la búsqueda, los caminos más transcurridos se hacen cada vez más probables hasta que al final todo el sistema converge en un único camino.

Puede ocurrir que el camino en el que se estabilizan las hormigas no sea el más corto. Pero así es cómo funciona la naturaleza. Demasiadas hormigas se morirían en la búsqueda si tuvieran que probar todos los caminos posibles.

En 1991 Dorigo⁷ desarrolló el primer algoritmo de optimización basada en colonias de hormigas. El algoritmo fue probado con un típico *problema del viajante del comercio (TPS traveling salesman problem)* que busca determinar cuál es el recorrido óptimo que un comerciante debe realizar para pasar por un conjunto de ciudades conectadas por distintas rutas con la mínima trayectoria posible.

La probabilidad con la que una hormiga elige el próximo nodo depende de dos variables: la distancia al nodo y la cantidad de feromona en el trayecto. Si sólo dependiera de la distancia, el sistema se estancaría rápidamente en mínimos locales pero gracias a la feromona que es depositada y luego evaporada hay probabilidad de que una hormiga elija un trayecto que sea más largo que otro si es que éste tiene suficiente feromona. Se permite al sistema que se empeore localmente para mejorar globalmente. De este modo se cuidan las dos características básicas de una optimización aproximada: la exploración y la explotación del espacio de soluciones.

⁷Dorigo M., V. Maniezzo & A. Colorni. **Positive Feedback as a Search Strategy**. Technical Report No. 91-016, Politecnico di Milano, Italy. 1991

El peso que se le da a cada variable es un parámetro que afectará la solución y dependerá de cada problema pero también puede optimizarse.

Desde su aparición en 1991 este algoritmo fue evolucionando y diferentes estrategias fueron probadas. Por ejemplo:

- *Ant Cycle*: cada vez que una hormiga termina un ciclo deposita feromona. La cantidad depositada es inversamente proporcional a la distancia recorrida. Acá se ve la primera diferencia con la realidad: las hormigas del modelo tienen memoria y recuerdan cuál fue el recorrido que hicieron.
- *Ant –density y ant-quantity*: la feromona es depositada cada vez que la hormiga elige un nodo.
- *Elistist strategy*: las hormigas salen en grupos. De cada grupo se elige la solución óptima y sólo ésta deposita feromona.
- *Max-Min Ant Systems (Stuzle and Hoos)*: la densidad de la feromona está limitada a un máximo y a un mínimo. La feromona depositada es máxima al principio de la optimización de modo que se potencia la exploración al principio y la explotación al final. Además utiliza la estrategia elitista.

2.4 Simulación y modelización

Dice la Real Academia Española: *Simular: Representar algo, fingiendo o imitando lo que no es.* Según R.E. Shannon⁸ *"La simulación es el proceso de diseñar un modelo de un sistema real y llevar a término experiencias con él, con la finalidad de comprender el comportamiento del sistema o evaluar nuevas estrategias -dentro de los límites impuestos por un cierto criterio o un conjunto de ellos - para el funcionamiento del sistema".*

En concreto, la simulación es una herramienta que se utiliza para entender, analizar, intervenir y tomar decisiones sobre un sistema. Es una excelente alternativa para sistemas que debido a su complejidad se vuelven difíciles de comprender y resulta imposible realizar conjeturas sobre su evolución frente a diferentes estímulos.

La fuerza de esta herramienta radica en que se puede entender cómo reaccionaría el sistema a una serie de cambios sin tener que efectuarlos en la realidad.

Gran parte de la efectividad de una simulación radica en el proceso de modelización, mediante el cual se construye un modelo de la porción de la realidad que se está analizando. Con el modelo no se pretende representar literalmente la realidad, sino que se busca entender y replicar las relaciones que lo afectan de manera suficientemente precisa como para poder estudiarlo y obtener conclusiones relevantes pero procurando mantener la simpleza.

⁸Shannon, Robert; Johannes, James D. (1976). «Systems simulation: the art and science». *IEEE Transactions on Systems, Man and Cybernetics*

2.4.1 El proceso de modelización

Se trata de un proceso de mejora continua basada en la retroalimentación. A continuación se explican las etapas involucradas en este proceso, basadas en una adaptación de Richardson y Pugh .

El disparador del proceso es la identificación del problema. Normalmente se comienza con una idea aproximada. Lo más probable es que una vez que se realiza el proceso de modelizado cambia la percepción del sistema real y por ende se re-define el problema.

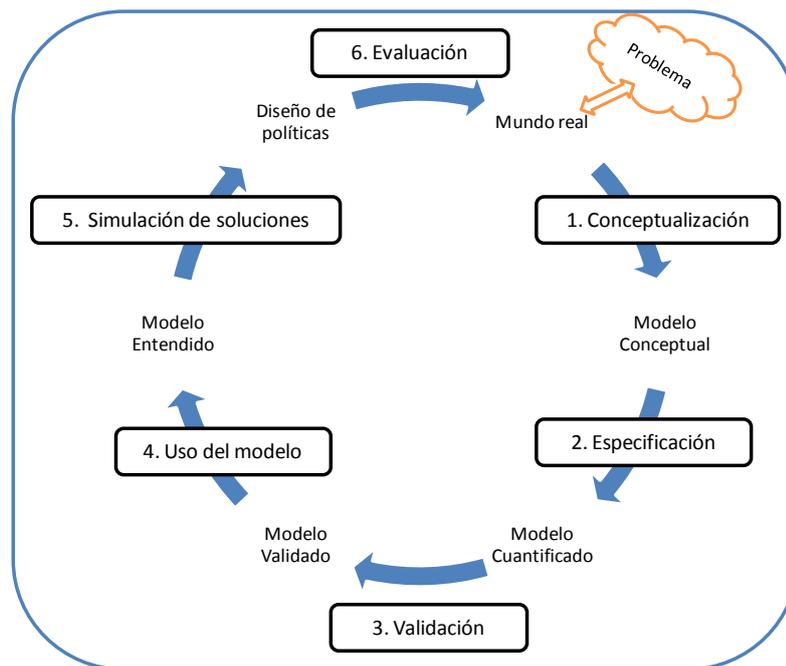


Diagrama 4: Proceso de Modelización. Adaptación de Introduction to System Dynamics and Modeling [Richardson; Pugh 1981]

1. **Conceptualización:** Se definen los límites del sistema, los cuales están directamente relacionados con la definición del problema y los alcances esperados del análisis. Luego se identifican los conceptos más importantes del sistema y se traducen en variables y relaciones cuantificables. De esta etapa surge el modelo conceptual que debe reflejar la estructura del sistema. El modelo conceptual puede ser validado pero únicamente en forma cualitativa.
2. **Especificación:** Es la cuantificación de las variables, y la definición de las relaciones. Se utilizan herramientas de análisis de datos como herramientas estadísticas y *data mining* para traducir la información del modelo conceptual al modelo cuantificado. Esta información se utiliza de input para el software de simulación. El output es un modelo que refleja la estructura del sistema y la evolución de las variables pero que todavía no está calibrado.
3. **Validación:** La validación del modelo es fundamental para que pueda tener utilidad práctica. Si bien la estructura sí está comprobada (por ejemplo si las relaciones son directas o inversas) falta calibrar las sensibilidades de las

relaciones. El método más común de calibración y validación es correr el modelo y comparar el resultado con la realidad. Por ejemplo, si se trata de la evolución de un sistema en el tiempo se corre el modelo en el pasado para poder comparar la salida del modelo con la realidad.

4. **Uso del modelo:** una vez que el modelo está validado se puede confiar en lo que arroja como salida. Con esto en mente se puede jugar con diferentes escenarios para entender cómo reaccionaría el sistema real. Se pueden buscar las situaciones de saturación, los escenarios radicales que deben ser evitados y las situaciones de mejora. Es cuestión de ordenarse en las pruebas y ser rigurosos en el análisis.
5. **Implementación:** De la etapa anterior se define cuál será la estrategia a seguir y se utiliza el modelo para entender cómo será la implementación de la misma en el mundo real. Se identifican cuáles serán los problemas potenciales y se desarrolla un sistema de medición y monitoreo de las variables.
6. **Evaluación:** una vez realizada la implementación, se analiza que el sistema real evolucione de la misma manera que el sistema modelado. Todo cambio en el sistema original, causará un nuevo “mundo real” de modo que el ciclo vuelve a comenzar.

2.4.2 Simulación por agentes

En la simulación por agentes se utilizan entidades que tienen reglas de comportamiento individuales. Cada entidad es un agente y tiene definido sus propios parámetros, reglas de decisión, y cualquier característica. Como consecuencia si se entiende cómo actúa cada agente individualmente y con su entorno, de manera simple puede generarse un sistema muy complejo.

Este tipo de simulación nos permite analizar el sistema en su totalidad de manera muy general, pero también puede analizarse hasta la entidad más pequeña del mismo. De modo que se puede ir desde lo más macro hasta lo más micro según sea necesario.

Algunos ejemplos de aplicación de simulación por agentes es: contagio de epidemias, modelos de optimización, estudios de bandadas de aves, evolución de redes sociales, etc.

3 CONTEXTO

Como ya fue explicado, el análisis de este trabajo está acotado a Google y a su servicio de publicidad: Google Adwords.

En la actualidad Google lidera el mercado de búsquedas en Internet con el 85% de ellas. Lo siguen Yahoo, Bing, Baidu (en China), Ask y muchos otros motores pequeños.

Esta cuota del mercado se traduce en cientos de millones de búsquedas diarias para Google. A continuación se ve la evolución de ganancias que implicó para Google la publicidad, desde su comienzo en 2001⁹.

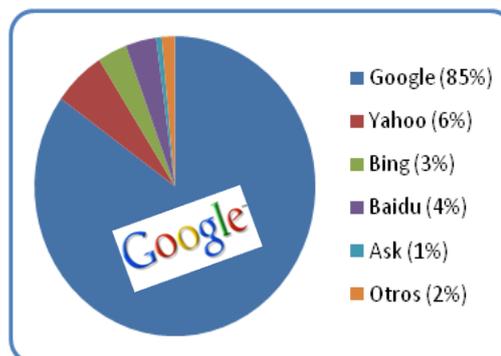


Gráfico 2: Market share promedio de los principales buscadores en internet (2010-2011)

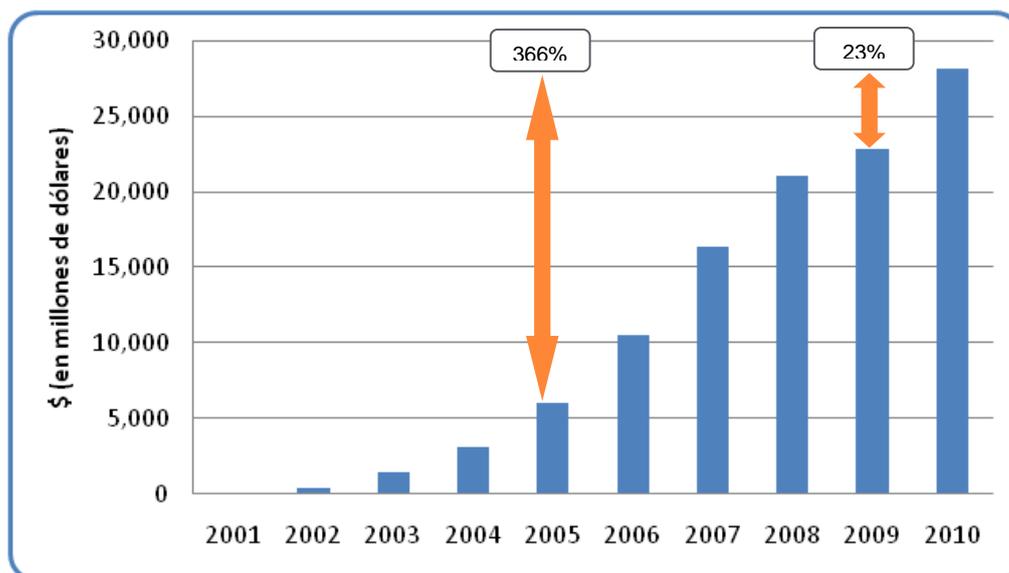


Gráfico 3: Ganancias de Google por publicidad en millones de dólares

Desde el año 2005 en adelante, las ganancias por publicidad implicaron en promedio el 98% de las ganancias totales de Google. En el año 2010 las ganancias de Google por publicidad fueron de 28,000 millones de dólares.

El aumento en ganancias de Google se debe a dos factores. Por un lado es un reflejo del aumento en cantidad de personas que utilizan Google Adwords como parte de su marketing. Por el otro lado se debe al aumento del costo promedio por *click*.

⁹ Fuente: <http://investor.google.com/financial/2003/tables.html> . Vigente al día 06/05/2011

Lo primero es un reflejo del aumento de relevancia que está cobrando la publicidad en internet con respecto a otros medios de publicidad más tradicionales.

El aumento del CPC (costo por *click*) es consecuencia del aumento de la competencia: a medida que aumenta el número de personas que subastan por la misma palabra, aumenta el precio que se debe pagar por ella. Por ejemplo, la frase “*flights Miami*” tiene un costo promedio por click de 2.20 dólares mientras que la frase “*Tampa to Jacksonville bus*” tiene un costo promedio por click de 0.41 dólares¹⁰. Ambas pertenecen a la misma industria pero por ser “*flights Miami*” un término más amplio que el otro, es mayor la oferta de subastas y por ende mayor el costo.

A modo de ejemplo se muestra a continuación la evolución del costo por click promedio para un grupo de 40 empresas de industrias variadas¹¹.

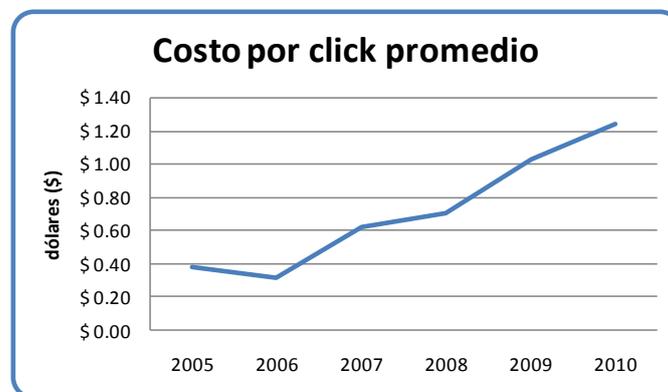


Gráfico 4: Evolución del costo por click promedio

El CPC promedio depende mucho de la industria en la que se esté publicitando. *Wordstreams* una empresa que vende software para analizar campañas publicitarias online. Según *Wordstream* la industria que más paga por sus *clicks* es la industria de los seguros, la cual genera más del 23% de los beneficios de Google. ¡El costo por *click* de palabras para esta industria llega a los 54 dólares! Es decir, una empresa está dispuesta pagar 54 dólares para que un potencial cliente realice un *click* sobre su anuncio. Otros ejemplos de las palabras más caras: préstamos (44 \$), abogado (42\$), rehabilitación (33 \$), etc.

En conclusión, durante los últimos años la industria de la publicidad en internet ha tenido un crecimiento significativo. Según un reporte del “*Internet Advertising Bureau (IAB¹²)*” este medio de publicidad creció un 16% en ganancias desde la primer mitad del año 2009 hasta la primer mitad del año 2010. Ningún otro canal de publicidad experimentó un cambio similar.

¹⁰Esta información fue sacada de una campaña actual para una empresa de transporte en Miami.

¹¹Esta información sale de HochmanConsultants <http://www.hochmanconsultants.com>

¹²Fuente: <http://www.iab.net>

4 METODOLOGÍA

A continuación se muestran cuáles son los pasos en la gestión de una campaña publicitaria.



Diagrama 5: Pasos a seguir en la gestión de una campaña publicitaria en Google Adwords

Una vez definido el presupuesto diario para toda la campaña se debe definir cuánto se está dispuesto a pagar por un *click* de cada una de las palabras que forman parte de la campaña.

Este estudio se centra en este último paso: **dado un grupo de palabras potenciales, y para un presupuesto diario dado, se compararán diferentes estrategias para la determinación de la subasta de cada palabra.**

4.1 Especificación del sistema teórico

Se llama sistema al conjunto de palabras potenciales¹³ y los atributos asociados a cada una de las posiciones potenciales. Cada palabra está asociada a un ingreso, un costo y un volumen que pueden variar con el tiempo y con la posición. Los atributos puede ser función de la palabra (kw^{14}), la posición (p) y el tiempo (t).

- Ingreso: aquello que el anunciante recibirá cuando un usuario realiza un *click*.
- Costo: aquello que el anunciante deberá abonar cuando un usuario realiza un *click*
- Volumen: la cantidad de *clicks* que esa palabra recibirá.

Las reglas de decisión son las que se utilizan para calcular la subasta máxima para cada palabra que dependerá de la estrategia utilizada. Una vez fijada la subasta máxima puede calcularse un resultado emergente.

Inputs del sistema

- **Presupuesto diario**

¹³Cuando decimos palabras potenciales, también nos referimos a un grupo de palabras o frases.

¹⁴KW por el nombre en inglés: *keyword*

- **Matriz de valores:** Hay una matriz por palabra definiendo para cada posición el costo, el ingreso y el número máximo potenciales de *clicks*.

Posición	Costo	Ingreso	Volumen
1	C_{i1}	I_{i1}	V_{i1}
2	C_{i2}	I_{i2}	V_{i2}
J	C_{iJ}	I_{iJ}	V_{iJ}

Tabla 3: Ejemplo de una matriz de valores para una palabra

- **Reglas de decisión:** Son la codificación de cada una de las estrategias. La aplicación de las reglas de decisión al sistema arroja como output un resultado.

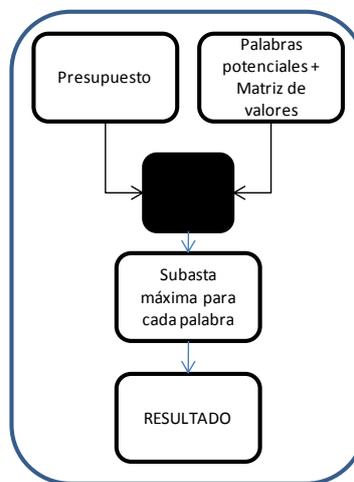


Diagrama 6: Inputs & Outputs del sistema

Outputs del sistema

Palabra	Posición	Subasta máxima
1	P_1	→ $\$1$
2	P_2	→ $\$2$
n	P_n	→ $\$3$

Diagrama 7: Ejemplo del output del sistema

Si bien en la realidad algunos de los atributos no son intrínsecos a la palabras sino que son dinámicos y dependen del conjunto de decisiones tomados por todos los actores involucrados, pueden realizarse ciertas simplificaciones para aproximarlos mediante funciones y así abordar este problema de forma teórica y llegar a conclusiones válidas.

A continuación se explican cada una de las simplificaciones, hipótesis y decisiones realizadas para llegar al sistema.

4.1.1 Costo

El costo asociado a una palabra depende tanto del tiempo como de la posición de la misma. Por definición del sistema de subastas el costo disminuye a medida que aumenta la posición.

Brendan Kitts y Benjamin Leblanc en su paper “Optimal Bidding on Keyword Auctions¹⁵” explican cómo se puede estimar la función que nos devuelve qué posición recibiríamos de subastar p \$ en el momento t. En el siguiente gráfico se

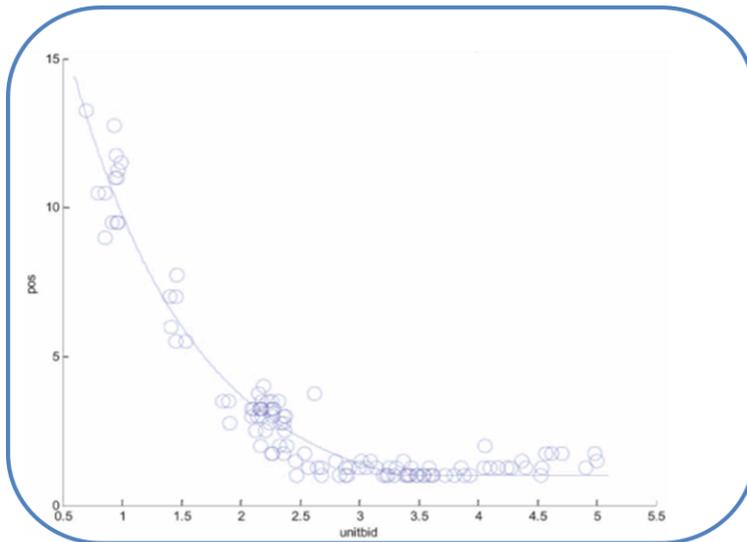


Gráfico 5: Gráfico extraído del paper “*Optimal Bidding on Keyword Auction*” de los autores Brendan Kitts y Benjamin Leblanc. Muestra la posición en función del precio.

ve una muestra de puntos y la función que minimizan su error. Para el ejemplo cualquier subasta mayor a 3.2 \$ resultará en la primer posición.

La función utilizada es una exponencial de parámetros θ y ω . Para cada palabra se deben buscar los parámetros que minimicen el error de la muestra de puntos. Se adoptan estos conocimientos para generar las curvas de costos para las diferentes palabras:

Según Kitts y Leblanc: la posición en función de la subasta máxima se puede ajustar con una función exponencial de parámetros θ y ω .

La curva que se genere está asociada a un período temporal denominado horizonte temporal de planeamiento. Si se trabaja con un horizonte diario los puntos serán de períodos diarios y las curvas emergentes servirán para un análisis diario. Para horizontes temporales menores, los autores explican cómo generar curvas de un período a partir de curvas de otro. Esto se escapa del análisis de este proyecto por lo cual no se entrará en más detalle.

Generación de la variable costo

Para cada corrida se determina la cantidad de palabras potenciales que se usarán. Para cada una de ellas se genera una curva de Costo siguiendo la hipótesis anterior y variando los parámetros de forma aleatoria. Las reglas utilizadas son:

$$\text{posición} = \theta e^{\text{precio}\omega}$$

¹⁵ Brendan Kitts, Benjamin Leblanc. *Optimal Bidding on Keyword Auctions*. *Electronic Markets*, 14(3):186-201, September 2004.

$$precio = \frac{\ln\left(\frac{posición}{\theta}\right)}{\omega}$$

Utilizando la información de la publicación previamente citada, y realizando un relevamiento de curvas reales se definen los rangos de los parámetros a utilizar:

θ aleatorio entre 15 y 45

ω aleatorio entre -0.9 y -0.1

Las funciones generadas varían para cada palabra pero siempre mantienen la forma. Por ejemplo para una corrida de 8 palabras se generaron las siguientes funciones de costo.

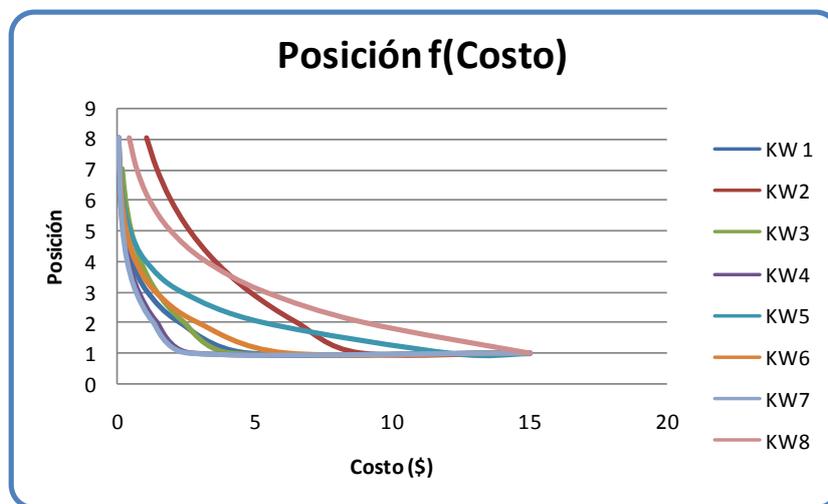


Gráfico 6: Generación del costo para la matriz de valores

A partir de estas funciones se puede identificar un costo para cada posición. El costo es el CPC que uno pagaría por esa palabra si la palabra apareciera en esa posición. Por ejemplo para la palabra 2 del gráfico 6:

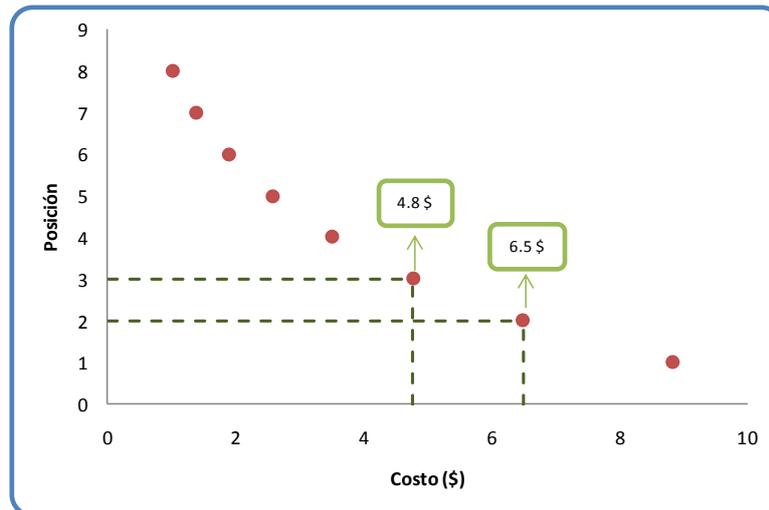


Gráfico 7: Curva costo-posición detallada para la palabra 2

Si la oferta máxima para esta palabra se encuentra entre 4,8 y 6,5 la palabra aparecería en la posición 3 y el CPC sería de 4.8.

Entonces a partir de estas curvas identificamos un CPC para cada posición. Este es el valor que se usa en la matriz de valores para el costo.

4.1.1 Ingreso

El ingreso asociado a una palabra depende de dos factores: las conversiones promedio por *click* y el ingreso promedio por conversión. Es decir de todas las personas que realizan un *click* en el anuncio y llegan a la página de destino, cuántas efectivamente realizan una compra. Además, ¿cuánto es el ingreso promedio por compra?

$$\text{Ingreso} = \text{ratio de conversiones} \times \overline{\text{Ingreso}}$$

Según Kitts y Leblanc¹⁶: el ingreso es independiente de la posición y del tiempo.

Esta hipótesis asume que una vez que un usuario llega a la página de destino tanto la probabilidad de que realice una conversión como el monto de la misma son independientes de la posición en la que se encontraba el anuncio cuando el usuario realizó el *click* como del momento del día y el día en que esto sucedió.

Generación de la variable ingreso

Para que el sistema teórico tenga sentido, se utiliza la información de la función de posición de la palabra para generar el ingreso.

¹⁶Brendan Kitts, Benjamin Leblanc. *Optimal Bidding on Keyword Auctions*. *Electronic Markets*, 14(3):186-201, September 2004.

- i. Se calcula el promedio del costo de todas las posiciones de la palabra
- ii. Se le suma un aleatorio al promedio (que puede ser positivo o negativo).

Como resultado de esta metodología en algunas palabras el beneficio por *click* (ingreso – costo) será negativo para algunas posiciones. Esto se hace para replicar lo que sucede a veces en la realidad: hay palabras que por tener mucha competencia tienen costos muy altos para las primeras posiciones. De nada sirve pagar ese precio si el ingreso asociado a esa posición para esa palabra no lo amerita. El ingreso dependerá de la industria en la que se esté, del nivel de calidad del anuncio, del porcentaje de conversiones asociadas a ese aviso y de muchas cosas; por eso se busca replicar este fenómeno en este sistema.

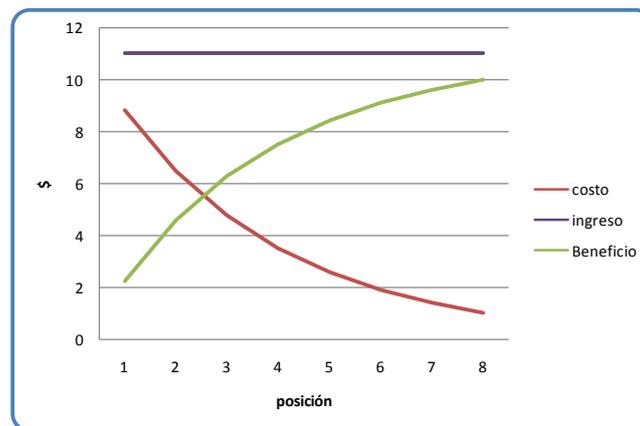


Gráfico 8: Para la palabra 2- Ingreso, Costo y Beneficio por click

En este ejemplo (nuevamente la palabra 2) el ingreso resultó ser mayor que el costo de todas las palabras. Queda claro que con un ingreso constante y un costo que disminuye se tiene un beneficio que aumenta a medida que aumentan las posiciones.

Entonces ¿por qué querríamos situar a las palabras en las primeras posiciones? La respuesta está en el volumen potencial de *clicks*.

4.1.2 Volumen

El volumen de *clicks* que recibirá una palabra también depende de la posición en la que se encuentra y del tiempo.

El igual que la función posición, la función del volumen que mejor ajusta a los puntos es una función exponencial de parámetros θ y ω .

Según Kitts y Leblanc¹⁷: la función del volumen se puede ajustar con una función exponencial de parámetros θ y ω .

$$\text{volumen} = \theta e^{pos\omega}$$

¹⁷Citar paper

De la misma manera que con el Costo, se relevan los parámetros para lograr curvas representativas de volumen.

⊖ aleatorio entre 5 y 15

ω aleatorio entre -0.9 y -0.3

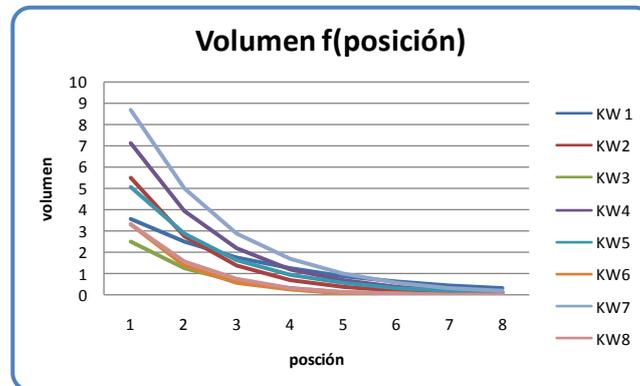


Gráfico 9: función volumen

Al igual que en el caso del costo, a partir de esta función se identifica cuál es el volumen de *clicks* que recibiría cada palabra si estuviera en cada posición (es decir si la subasta máxima fuera mayor al costo calculado con la función posición).

Además se utiliza la siguiente **hipótesis: los *clicks* de cada palabra están distribuidos uniformemente en el horizonte de planeamiento.** Esta consideración se utiliza para ajustar el resultado al presupuesto diario (4.14)

4.1.3 Presupuesto

Si bien el presupuesto es un dato que se le agrega al sistema como input según cuánto dinero quiera gastar quien maneja la cuenta, en este sistema teórico se genera un presupuesto siguiendo una lógica para que el resultado tenga sentido.

Se generan deliberadamente tres situaciones:

- i. Un presupuesto/ palabra alto
- ii. Un presupuesto/ palabra medio
- iii. Un presupuesto / palabra bajo

Para cada corrida se elige aleatoriamente una de las tres situaciones y luego se define el presupuesto multiplicando la cantidad de palabras por un factor cuyo valor depende de la situación a la que pertenece la corrida.

4.1.4 Cálculos

$$\text{Beneficio} = \text{Ingreso} - \text{Costo}$$

$$Beneficio_{por\ click} = Ingreso_{por\ click} - Costo_{por\ click}$$

$$Beneficio_{total} = Beneficio_{por\ click} * Volumen$$

$$Beneficio_{por\ \$} = \frac{Ingreso - Costo}{Costo}$$

4.1.5 Ajuste del resultado al presupuesto

En la realidad, cuando una campaña cumple su presupuesto diario dejan de imprimirse las palabras y como resultado inmediatamente la campaña deja de recibir *clicks*. Esto se replica en este sistema teórico de la siguiente forma:

Se calculan los resultados del sistema (costo, ingreso, volumen) según la estrategia planteada. El volumen que se usa es el potencial es decir el volumen de *clicks* que recibiría la palabra si se imprimiera durante todo el período. Si dicho resultado implica un costo que supera el presupuesto diario planeado se ajustan los resultados de manera lineal:

$$Beneficio\ ajustado = Beneficio\ base * \frac{presupuesto}{Costo\ base}$$

Esto se cumple bajo la hipótesis de que la distribución de *clicks* de todas las palabras es uniforme en el horizonte temporal de planeamiento. Cuanto menor sea el horizonte más se ajustará la hipótesis a la realidad.

4.1.6 Índice de Calidad

Se vio en el marco teórico el efecto del índice de calidad en la subasta. Se vio que el precio que pagará un anunciante equivale al producto de la oferta máxima del siguiente competidor y su nivel de calidad sobre el nivel de calidad propio.

$$P_1 C_1 = O_2 C_2$$

$$P_1 = \frac{O_2 C_2}{C_1} = \frac{Ad_2}{C_1} \rightarrow (Ad_2: AdRank_2)$$

De esta ecuación se desprende la siguiente conclusión: **si un anunciante mejora su nivel de calidad, y la competencia se mantiene igual, su precio disminuirá.**

A modo de ejemplo se presenta la siguiente tabla. Supongamos que un anunciante tiene un índice de calidad = 6. Si el índice mejora a 7, su precio se verá reducido en un 14%. Si por el contrario su índice disminuye a 5 su precio se verá incrementado un 20%.

CPC vs un nivel de calidad = 6		
Nivel de calidad		CPC
10	-	40%
9	-	33%
8	-	25%
7	-	14%
6		0%
5	+	20%
4	+	50%
3	+	100%
2	+	200%
1	+	500%

Tabla 4: Efecto del cambio del nivel de Calidad en el CPC¹⁸

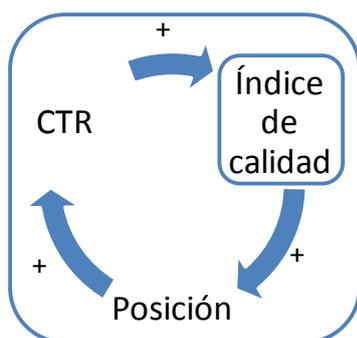


Diagrama 8: Lazo reforzador para el Índice de calidad

Por otro lado, se sabe que el CTR (%clicks/impresiones) aumenta a medida que aumenta la posición. Y asimismo a mayor CTR mayor índice de calidad.

De modo que se crea un lazo reforzador del cual queda claro que al anunciante le conviene mejorar su índice de calidad lo más posible. Para hacerlo deberá invertir tiempo y dinero, es por eso que puede haber

un límite que esté dispuesto a pagar.

El efecto de una mejora en el índice de calidad es tanto mayor que el efecto de una mejora en la manera de subastar, que no tiene sentido optimizar las subastas de palabras si el índice de calidad no es tan bueno como podría serlo.

De modo que este análisis asume que el índice de calidad ya está estabilizado y es constante, y cuando se definen las curvas de costos y volumen ya está implícito el efecto del mismo.

4.1.7 Horizonte temporal de planeamiento

Llamamos horizonte temporal de planeamiento al período de tiempo contemplado en la comparación de estrategias. El horizonte temporal de este análisis es de un

¹⁸Idea adaptada de *Click Equations INC.* <http://www.clickequations.com/>

día. Es decir, se toman las decisiones una vez al día y la matriz de valores para cada palabra rige a lo largo del mismo.

Para los alcances de este estudio alcanza con utilizar un horizonte temporal de planeamiento diario porque es suficiente para comparar las diferentes estrategias. Las conclusiones luego podrán ser aplicadas para diferentes horizontes temporales manteniendo la metodología y adecuándose los inputs. Simplemente cambiaría un poco la estructura de algunas de las estrategias, pero eso se explicará en el capítulo 6 en las futuras investigaciones.

4.2 Identificación de las estrategias

Ahora que el sistema teórico ha sido detallado es más fácil comprender dónde radica la complejidad del problema puntual de la definición de subasta máxima. El costo disminuye de manera no lineal a medida que aumentan las posiciones haciendo más atractivas las posiciones más altas. El volumen también disminuye haciendo más atractivas las primeras posiciones. Entonces hay que encontrar el balance óptimo entre posiciones de margen alto y volumen bajos y las de margen y volumen alto.

Posiblemente existan tantas estrategias como personas que manejan cuentas de AdWords. Hay quienes siguen estrategias definidas y consistentes en el tiempo pero hay también quienes toman decisiones que no son reproducibles de manera automática por utilizar factores no cuantificables como su intuición o consejos externos.

Se buscan para este análisis estrategias que sean **automáticas, consistentes, extrapolables y reproducibles**. Estas cualidades son necesarias para poder llevar las conclusiones de un análisis a horizontes más amplios sin estar restringidos por la cantidad de palabras, el horizonte temporal, el presupuesto diario u otras variables del sistema.

Las estrategias planteadas a continuación surgieron en parte de la literatura escrita sobre este tema y en parte de hablar con personas con experiencia y conocimiento. El paper previamente citado de los autores Kitts y Leblanc con un enfoque innovativo propone resolver el tema desde el punto de vista de una optimización. Sus ideas e hipótesis son utilizadas y adaptas para una de las siguientes estrategias.

Tipo 1: Estrategias equitativas

Este grupo de estrategias busca equilibrar algún indicador en cada una de las palabras. Si la solución implica un costo mayor al presupuesto, se ajusta el resultado. El indicador utilizado es la subasta máxima equitativa.

1.a Subasta máxima equitativa

Esta estrategia reparte equitativamente el presupuesto entre las diferentes palabras potenciales.

Metodología

1. Se fija la subasta máxima para cada palabra

$$subasta = \frac{presupuesto}{cantidad\ de\ palabras}$$

2. Para cada palabra se calcula la posición correspondiente a esa subasta:

$$posición = \frac{presupuesto}{cantidad\ de\ palabras}$$

3. Se calcula el Beneficio para esa posición.

4. Si el beneficio es negativo para alguna palabra, se descarta la misma y se vuelve a calcular la máxima subasta por palabra para las que quedan.

5. Se repiten los pasos 2-3-4.

6. Se calcula el resultado, si es necesario se ajusta.

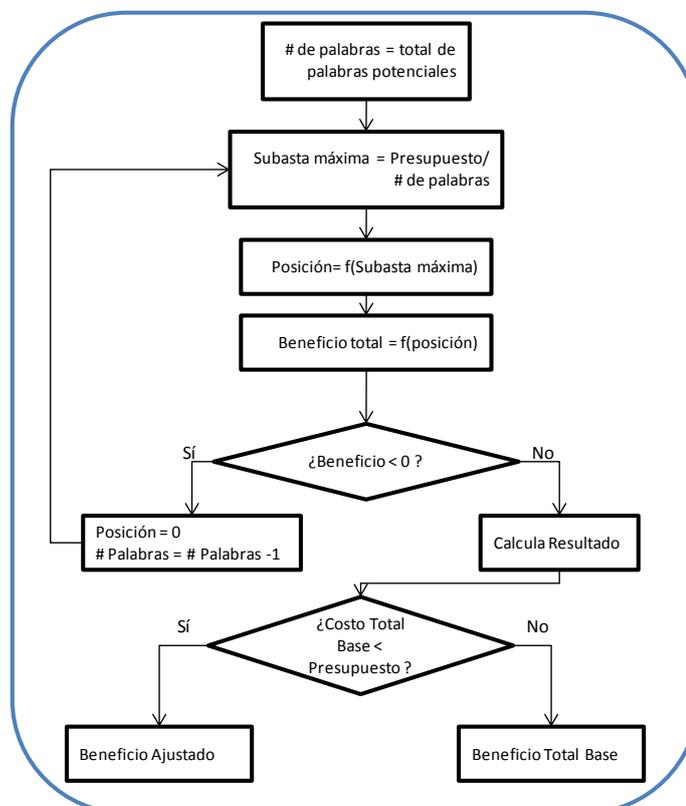


Diagrama 9: Estrategia 1

El único discernimiento que se efectúa entre palabras es la eliminación de aquellas que arrojan un beneficio negativo para la posición designada. Claramente si el portfolio de palabras potenciales se viera reducido, cambiaría el resultado de esta estrategia. De modo que podría limitarse la cantidad de palabras potenciales con algún criterio especificado antes de utilizar esta estrategia.

Tipo 2: Estrategias de maximización por palabra

Este grupo de estrategias busca maximizar algún indicador para cada una de las palabras. Si la solución implica un costo mayor al presupuesto, se ajusta el resultado.

Metodología

1. Se calcula el indicador para cada posición de cada palabra= I_{ij}
2. Se identifica la posición que maximiza el indicador para cada palabra $j_{i\max}$
3. Se calcula el costo total base y el beneficio total base que implican dichas posiciones.
4. Si el costo total base excede el presupuesto se ajusta el beneficio total.

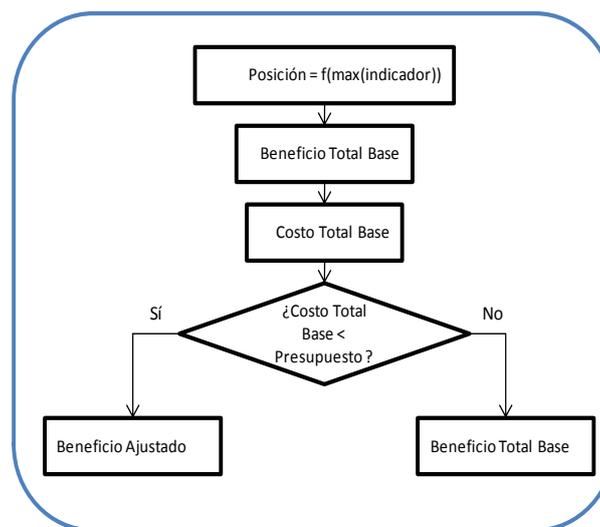


Diagrama 10: Estrategia tipo 2

Esta estrategia se prueba con tres indicadores diferentes: beneficio total, beneficio por \$ invertido por *click* (beneficio total / costo total) y beneficio por \$ invertido total (beneficio total/ costo total) x volumen.

2.a. Maximización del beneficio total

Indicador = beneficio total = (ingreso – costo) x volumen

Para cada palabra se busca la posición que maximice el beneficio total. Los valores unitarios graficados de ejemplo, mantienen las formas de las curvas previamente explicadas:

- Ingreso: constante
- Costo y Volumen: descienden exponencialmente.

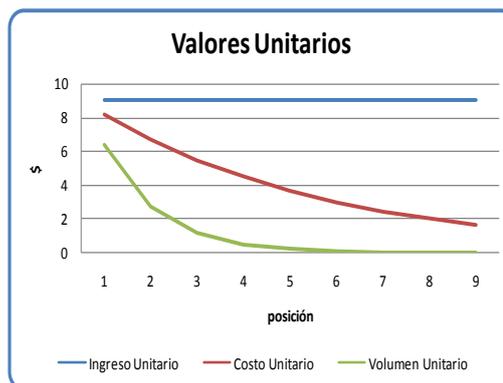


Diagrama 11: Valores Unitarios

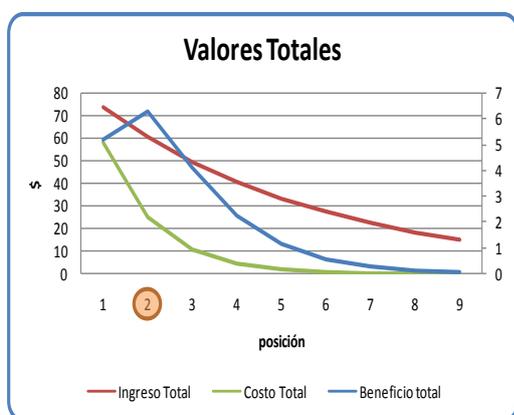


Diagrama 12: Valores totales

Estos valores unitarios conllevan los siguientes valores totales:

En este ejemplo la posición 2 maximiza el indicador.

Esta estrategia determina la posición para cada palabra y luego ajusta el resultado final.

Este enfoque maximiza el ingreso total por palabra, pero no contempla la eficiencia con la que se invierte cada unidad de dinero.

2.b. Maximización del beneficio por \$ invertido

Indicador = (Ingreso – Costo)/Costo

Contrario al caso anterior, este enfoque contempla únicamente la eficiencia con la que se invierte cada unidad de dinero maximizando el retorno por cada unidad invertida pero sin contemplar el volumen implicado. Como el ingreso es constante y el costo disminuye, el beneficio por \$ aumenta a medida que aumenta la posición.



Diagrama 13: Beneficio/ \$

Lógicamente la última posición siempre maximizará el indicador. El riesgo de esta estrategia está en que las últimas posiciones pueden tener volúmenes tan bajos que el resultado final termina estando muy lejos del óptimo.

2.c. Maximización del (beneficio por \$ invertido)x volumen

Indicador = $(\text{Ingreso} - \text{Costo})/\text{Costo} \times \text{volumen}$

Este enfoque busca contemplar ambas cosas al mismo tiempo: se trata de invertir eficientemente mirando al mismo tiempo el volumen potencial. Ambos factores pesan lo mismo de modo que podría ocurrir cualquiera de los dos casos anteriores. Para el mismo ejemplo nuevamente gana la posición 2.

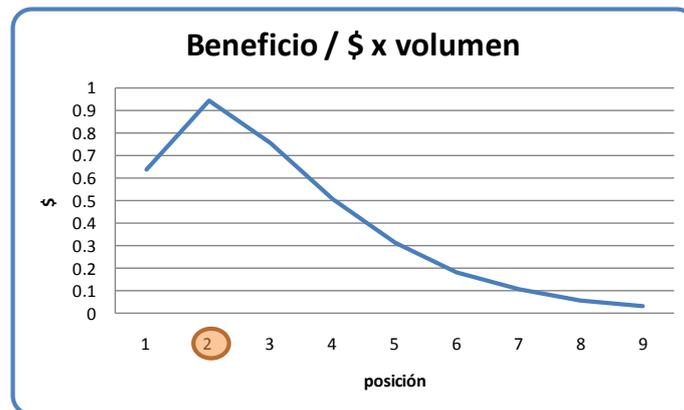


Gráfico 10: beneficio/ \$ x volumen

Tipo 3: Estrategias de ranking

Este grupo de estrategias también busca maximizar un indicador para cada palabra, pero utiliza tantas palabras como le convenga al sistema.

Metodología

1. Se calcula el indicador para cada posición de cada palabra= I_{ij}
2. Se identifica el indicador máximo para cada palabra, y la posición que lo maximiza I_{maxi}, j_{maxi}
3. Se realiza un ranking de indicadores máximos, ordenados de mayor a menor
4. Siguiendo el orden del ranking se agrega de a una palabra a la vez hasta que agregar una palabra más empeore el resultado final. Al resto de las palabras se la asigna la posición 0 y luego se ajusta el resultado.

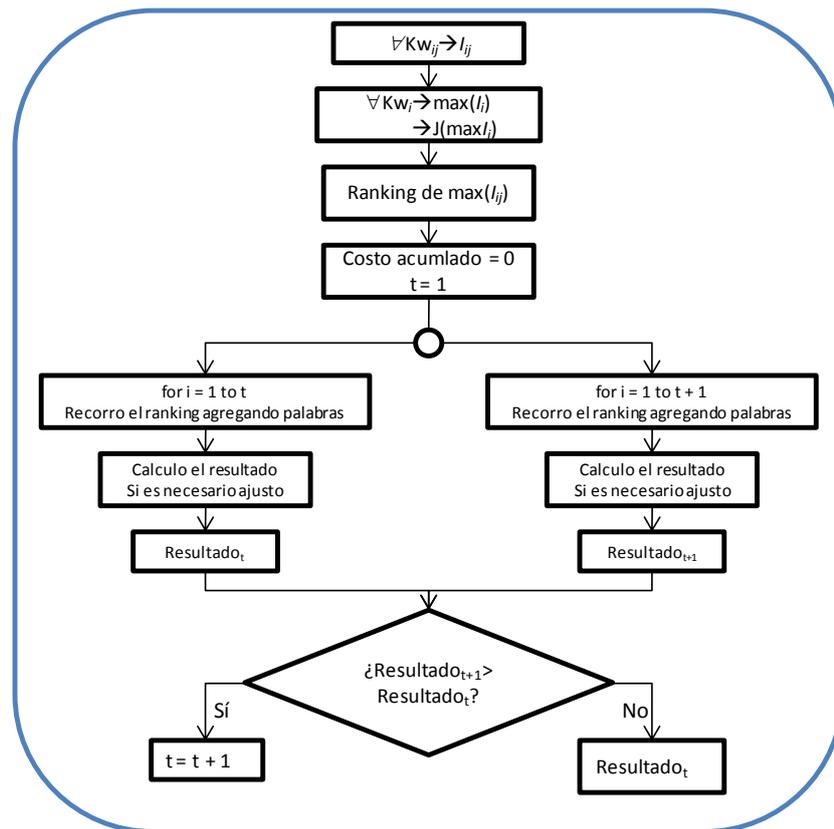


Diagrama 14: Estrategia Tipo 3

Este enfoque tiene un grado más de integración que el enfoque anterior. Ya no optimiza palabra por palabra sino que intenta mirar el sistema en conjunto, aunque parcialmente.

La idea es impedir que las palabras con bajo rendimiento empeoren el resultado final. Como no hay una única manera de medir este rendimiento se vuelven a probar tres indicadores diferentes:

3.a. Ranking de beneficio total

Indicador = (ingreso – costo) x volumen

3.b. Ranking de beneficio/\$

Indicador = (ingreso – costo)/costo

3.c. Ranking de beneficio/\$* volumen

Indicador = (ingreso – costo)/costo x volumen

Tipo 4: Estrategias de optimización global

Este grupo de estrategias intenta resolver el problema buscando un resultado óptimo global. Mira al sistema integralmente buscando evitar óptimos locales en pos de encontrar un máximo global.

Se utilizan dos métodos de optimización diferentes: ACO: optimización basada en colonias de hormigas y Método exhaustivo de optimización.

4.a ACO: optimización basada en colonias de hormigas

Esta optimización utiliza una adaptación del método de optimización basada en colonias de hormigas que se aplica mediante una simulación por agentes.

Este algoritmo de optimización fue creado basado en la manera en que las hormigas recorren espacios de soluciones para encontrar un camino óptimo entre su nido y la fuente de comida. La probabilidad con la que una hormiga elige un camino es función de la cantidad de feromona que siente y de la visibilidad del próximo punto. El algoritmo fue probado y utilizado para optimizar problemas de minimización de distancias adaptando el comportamiento real de las hormigas al problema para hacerlo todavía más efectivo.

Para resolver esta optimización se crea un paralelismo entre la minimización de la distancia y la maximización del beneficio. Se utiliza el mismo principio que en la optimización base pero ya no se busca un recorrido físico; se adecua el concepto de recorrido a la combinación de palabras y posiciones.

La optimización se lleva a cabo mediante una simulación por agentes. Como fue explicado en el capítulo 2, la simulación por agentes se basa en entidades que toman decisiones individuales a partir de las cuales emerge un comportamiento global del sistema. El comportamiento emergente de este modelo es el descubrimiento de nuevas combinaciones de palabras y posiciones que pueden ser las óptimas.

Estructura del modelo de simulación para la optimización

El agente principal de este modelo es el agente **hormiga**. Cada hormiga nace, "recorre un camino" (recorre las diferentes palabras eligiendo sus posiciones probabilísticamente), al hacerlo deposita feromona y luego muere.

Las hormigas nacen de a **lotes**. Cada lote de hormigas conforma un **ciclo**.

Además hay un agente que se llama **palabra** que tiene tantas replicaciones como potenciales palabras. Y un agente que se llama **posición** que tiene tantas replicaciones como posiciones posibles + 1. Se agrega una posición que es la posición 0; si una palabra está en esta posición no saldrá impresa y por ende no podrá recibir *clicks*.

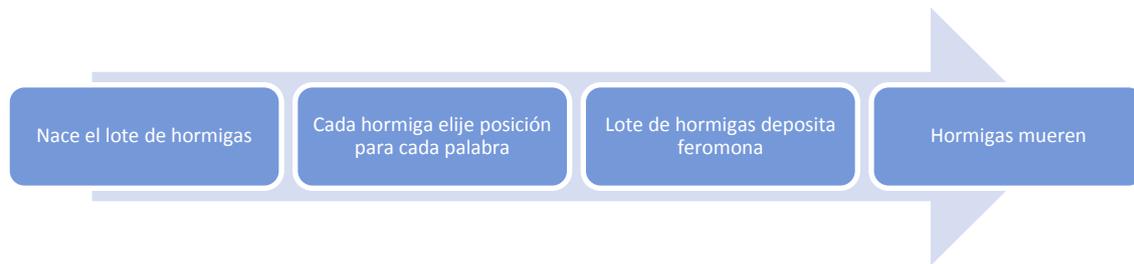
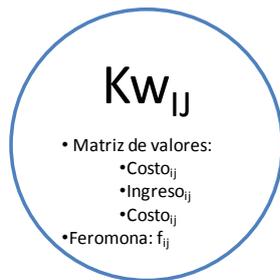
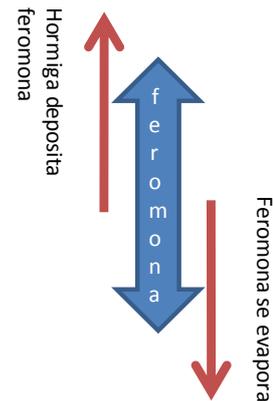


Diagrama 15: evolución de cada ciclo de hormigas



Para cada posición de cada palabra está definida la matriz de valores que es fija. Además en cada momento de la simulación está definida la cantidad de feromona. La feromona varía a lo largo de la simulación en función de cuántas hormigas pasaron por esa posición y de

la evaporación.



El sistema simulado evoluciona con la interacción de ambas partes. A medida que avanza la simulación y progresan los ciclos el sistema se estabiliza en una solución.

La simulación tiene una serie de parámetros que afectarán al sistema y podrán incidir en la solución en la que se estabiliza la optimización. A continuación se explica en detalle el funcionamiento del modelo destacando los principales parámetros a definir.

Metodología

Se tiene un sistema conformado por I palabras potenciales y (J+1) posiciones posibles para cada palabra. Se conoce la matriz de valores para cada palabra KW_{ij} y se conoce el presupuesto diario disponible.

1. **El modelo de simulación comienza con el armado de la estructura del mismo.** Se crean y replican los agentes, se importan los valores y se calcula para cada K_{ij} : beneficio total, costo total y beneficio total ajustado¹⁹.

$$beneficio\ total\ ajustado = \begin{cases} 0, & b < 0 \\ b + m, & b \geq 0 \end{cases}$$

$$b = beneficio\ total$$

$$m = \min(b) \forall b \geq 0$$

¹⁹Explicación del beneficio ajustado en el punto 3

2. **Cada posición de cada palabra comienza con una feromona inicial f_0** (incluyendo la posición 0). De modo que el sistema comienza con una *feromona inicial* $= f_0 \times I \times (J + 1)$
3. **Se calcula el atractivo de cada posición para cada palabra.** El atractivo es una función que está compuesta por los dos atractivo parciales: el atractivo por la feromona y el atractivo por el beneficio total. El peso de cada atractivo parcial es un parámetro fijo del sistema que está definido tal que entre ambos sumen 1.

$$\begin{aligned} \text{Atractivo} &= \sqrt{\text{atractivo feromona}^2 + \text{atractivo beneficio total}^2} \\ \text{Atractivo} &= \sqrt{(\text{peso } f \times \text{componente } f)^2 + (\text{peso } b \times \text{componente } b)^2} \\ \text{Componente feromona} &= \frac{f_{ij}}{\sum_{j=0}^{J+1} f_{ij}} \\ \text{Componente beneficio} &= \frac{B_{ij}}{\sum_{j=0}^{J+1} B_{ij}} \\ \text{peso feromona } p_f + \text{peso beneficio } p_b &= 1 \end{aligned}$$

En realidad el cálculo del atractivo utiliza el **beneficio total ajustado** como el segundo componente. La razón por la cual se ajusta el beneficio total es la siguiente:

Lógicamente la solución óptima puede requerir que una palabra aparezca en la posición 0, es decir que nunca aparezca impresa. Para permitir que esto suceda en la optimización debe haber una cierta probabilidad de que la hormiga elija esta opción. Se ajusta la variable beneficio total para aumentar dicha probabilidad.

Se realizan dos ajustes diferentes:

- Si el beneficio total de una posición es negativo, se pone = 0. Si lo que se busca es optimizar el beneficio total, esta opción no va a convenir nunca.
- Si el beneficio es mayor o igual a 0 se le suma una cantidad constante. Esta cantidad será igual al beneficio total mínimo de todas las posiciones con beneficio total mayor a 0. Se busca de este modo aumentar el atractivo de la posición 0 manteniendo el orden inicial de atractivo del resto de las posiciones.

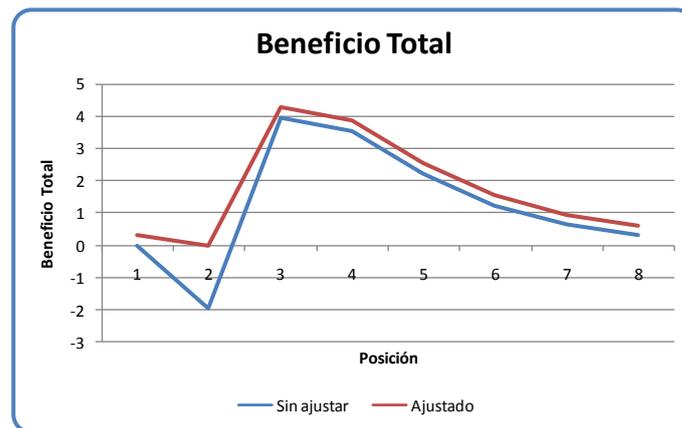


Gráfico 11: ajuste del beneficio total

4. **Nace el primer lote de hormigas.** Para cada hormiga del lote de hormigas el atractivo de cada posición es fijo. Cada hormiga recorre cada una de las palabras y elige la posición mediante una función de probabilidad uniforme, que depende linealmente del atractivo calculado.
5. Una vez que todas las hormigas del lote terminan este proceso **se deposita la feromona**. Al mismo tiempo se ejecuta la función que **evapora la feromona** del ciclo anterior.

La evaporación de la feromona es proporcional a la cantidad que hay: siempre se evapora un porcentaje f_e . Como la feromona total es contante, la cantidad de feromona evaporada por ciclo será constante.

$$\text{Feromona inicial por palabra} = (J+1) f_0$$

$$\text{Feromona inicial del sistema} = (J+1) f_0 I$$

$$\text{Feromona evaporada por ciclo} = (J+1) f_0 I \times f_e$$

Cada hormiga deposita la misma cantidad de feromona en cada uno de los puntos de su recorrido. Cuánto deposita depende del resultado de su combinación de modo que depositará más feromona la hormiga cuya combinación implique el resultado óptimo del ciclo.

$$\text{Feromona depositada por hormiga} = \text{Factor}_{\text{depósito}} \times \text{Factor}_{\text{resultado}}$$

$$\text{Factor}_{\text{resultado}} = F_r = \frac{\text{beneficio}_{\text{ajustado}}}{\text{beneficio}_{\text{total del ciclo}}} = \frac{b_a}{b_t}$$

$$\sum_{i=1}^{kw} F_{ri} = 1$$

$$\text{feromona depositada por ciclo} = \sum f_r \times f_d \times I = f_d \times I$$

Como la feromona total del sistema se mantiene constante en todo momento:

$$\text{Feromona depositada} = \text{Feromona evaporada}$$

$$f_d \times I = f_e \times I \times f_0 \times (J+1)$$

$$f_d \times I = f_e \times I \times f_0 \times (J+1)$$

$$f_d = f_e \times f_0 \times (J + 1)$$

$$\text{feromona depositada por hormiga} = f_e \times f_0 \times (J + 1) \times \frac{b_a}{b_t}$$

6. Se identifica y se guarda el resultado óptimo del lote.
7. Se compara contra el resultado óptimo global
8. Finaliza el modelo

Interface del modelo

El panel de control del modelo se utiliza para definir la cantidad de palabras, cantidad de posiciones y el presupuesto diario de la corrida. Una vez que comienza la corrida el modelo importa la matriz de valores de cada palabra desde un archivo Excel previamente generado y aparece en la



Diagrama 16: Panel de control del modelo

pantalla la siguiente interface.

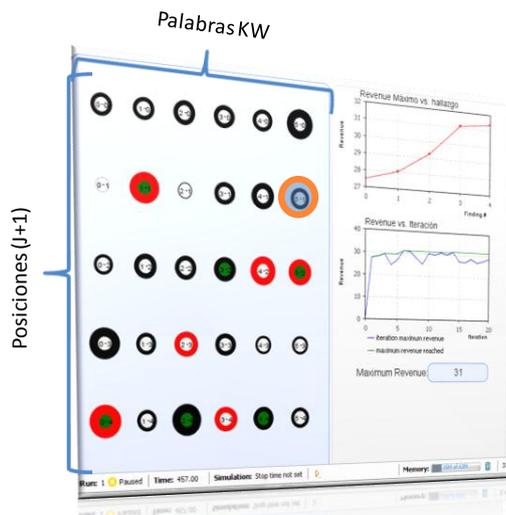


Diagrama 17: Interfase del modelo mientras corre

La pantalla está compuesta por dos partes: en la izquierda están esquematizadas todas las palabras en todas las posiciones. En la derecha hay gráficos que muestran la evolución de la corrida.

El punto azul representa la palabra 6 en la posición 1 (la primer fila corresponde a la posición 0 es decir es la situación en la que la palabra no aparece impresa).

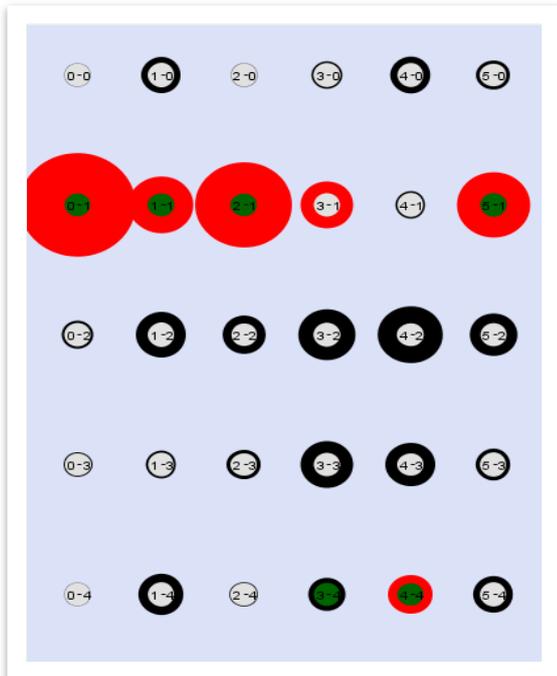


Diagrama 18: Interface de la mitad de la corrida

A la derecha de la pantalla aparecen dos gráficos. El de arriba muestra cada nuevo óptimo que se encontró. El de abajo tiene dos líneas. La línea azul muestra el óptimo de cada lote de hormigas y la línea verde muestra el óptimo global hasta el momento que es el mismo que aparece debajo: 103 \$ para el ejemplo.

El grosor del borde de cada diagrama es proporcional a la feromona que tiene la palabra en ese momento. Las palabras que están pintadas de verde son aquellas que forman el recorrido óptimo de ese lote de hormigas. Las palabras que tienen el borde rojo son aquellas que forman el recorrido óptimo global (es decir de todos los lotes de hormigas que ya pasaron, ese fue el mejor recorrido hasta el momento de la corrida).

A medida que transcurre la corrida suele suceder que la cantidad de feromona crezca mucho para ciertas palabras. Pero el sistema siempre mantiene la cantidad de feromona total constante.

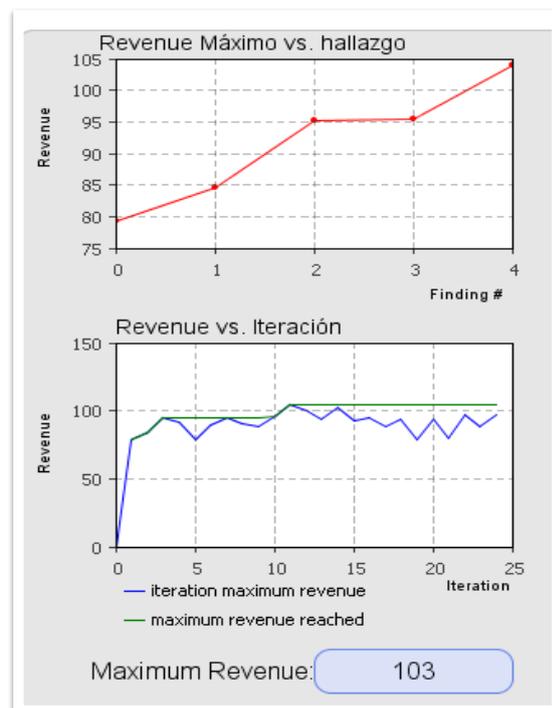


Diagrama 19: Gráficos de la corrida

Validación del modelo

Validar un modelo es imprescindible para poder confiar en el mismo y en las conclusiones que este arroja. De nada sirve obtener resultados óptimos en la simulación si luego no aplican en la realidad.

Como el único objetivo de este modelo es optimizar un problema de combinaciones, la manera de validar si la herramienta funciona correctamente es comprobar si el resultado de la optimización coincide con el óptimo real del sistema²⁰.

El proceso de validación se separó en dos fases: la determinación y optimización de los parámetros del modelo y la validación del output.

1. Determinación y optimización de los parámetros del modelo

Los parámetros a definir son

- **Cantidad de hormigas en cada lote**

Este parámetro es función de la cantidad de posiciones que tiene la corrida. Hormigas = (pos +1) x 3. Este número se determinó probando diferentes opciones. Si el número de hormigas disminuye mucho aumenta la cantidad de ciclos necesarios para encontrar el óptimo. Si por el otro lado el número de hormigas aumenta mucho, el modelo pierde efectividad. A mayor cantidad de hormigas por ciclo, menor es el peso relativo de cada recorrido. Si aumenta mucho el número de hormigas aumenta el porcentaje de veces que el modelo no encuentra el óptimo.

- **Peso de la feromona y peso del beneficio total (1 – p_f)**

Este es el parámetro más difícil de definir porque el óptimo posiblemente varíe para cada corrida en función de la matriz de valores. Además es probable que el óptimo sea dinámico y varíe a lo largo de la corrida. Se realizaron muchas corridas y se encontró que el óptimo del peso de la feromona variaba aproximadamente entre 90% y 100%. La mejor solución encontrada fue utilizar una función sinusoidal para que este parámetro tome esos valores a lo largo de la corrida.

$$peso_{feromona} = 0.9 + 0.1 * abs(\sin(2 * Pi * x))$$

$$x = f(ciclo)$$

$$peso_{beneficio} = 1 - peso_{feromona}$$

²⁰Igualmente para que el modelo esté validado no se exige que el modelo llegue al óptimo el 100% de las veces, simplemente porque los métodos heurísticos de optimización funcionan con esa premisa: nunca aseguran con certeza que funcionan encontrando el óptimo.

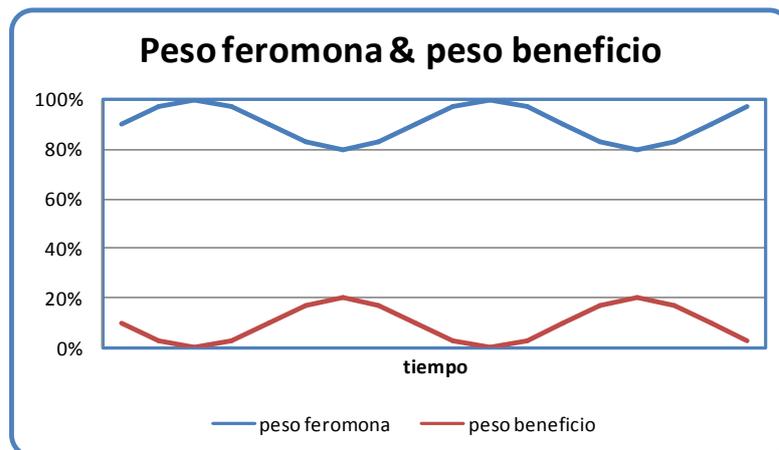


Gráfico 12: Pesos (tiempo)

• Factor de desgaste & Feromona por hormiga

Ambos parámetros fueron optimizados conjuntamente porque son dependientes. El software utilizado para crear el modelo de simulación (Anylogic) tiene un optimizador incorporado que se llama *Optquest*. El usuario debe definir cuáles son los parámetros a optimizar, fijando los límites posibles y las restricciones si las hay. Luego se corre el modelo de simulación y de manera eficiente el programa varía los parámetros hasta encontrar el óptimo. Cuando se trata de corridas estocásticas, como en este caso, se realizan diferentes corridas con una misma combinación de parámetros.

Los parámetros a variar fueron:

- Factor de desgaste (se varió entre 0.05 y 0.5 de forma discreta con un escalón de 0.05).
- Feromona por hormiga (se varió entre 1 y 10 de forma discreta con un escalón de 1).

Para diferentes escenarios (variando la cantidad de palabras, cantidad de posiciones y matriz de valores) se encuentra el óptimo mediante la optimización exhaustiva. Luego se le agrega al modelo ACO una función que se ejecuta cuando encuentra el óptimo. La función realiza dos cosas:

- Identifica el tiempo real que ha tardado en encontrar el óptimo, y llena la variable "optimizar" con este tiempo.
- Finaliza el modelo

```
if (nuevo_optimo == optimo_exhaustivo)
    {optimizar = getEngine().getRunTimeMillis();
    getEngine().finish();
};
```

Además si transcurren 2 minutos²¹ y la corrida no llega al óptimo se finaliza el modelo.

La variable **optimizares** la variable objetivo de la optimización. Se busca minimizarla. El valor inicial para la misma es de 120000 milisegundos (2 minutos). Si en una corrida no se encuentra el óptimo la variable permanecerá con ese valor.

Esta optimización se corrió diferentes veces. A continuación se muestra una de estas corridas.

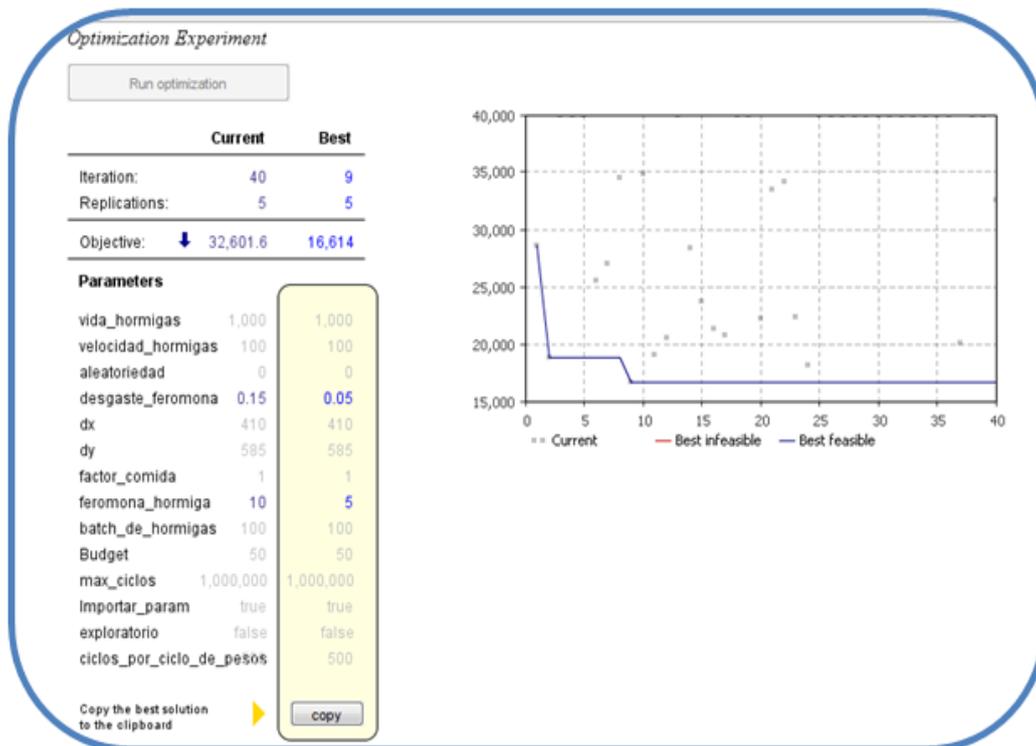


Diagrama 20: Ejemplo de optimización

Se realizan 5 replicaciones por cada combinación de parámetros por tratarse de un modelo estocástico. Cada uno de los puntos del gráfico de la derecha es el mejor resultado de las 5 replicaciones. Cada vez que uno de estos puntos significó un nuevo óptimo, pasó a formar parte de la línea azul. La combinación óptima para esta corrida fue: factor de desgaste = 0.05 y feromona por hormiga = 5.

Se corrió la optimización para diferentes escenarios. En todos los escenarios el factor de desgaste óptimo fue de 0.05. Sin embargo la cantidad óptima de feromona por hormiga varió mucho.

La conclusión es que la sensibilidad de la **feromona por hormiga** es muy baja. Se utiliza **5** por ser el que más veces salió en las optimizaciones. El **factor de desgaste** tiene más efecto en el resultado y se usa **0.05**.

²¹Justo a continuación se explica cómo se define la duración del modelo.

• **Tiempo**

Otro parámetro a definir es cuánto tiempo corre el modelo. Por tratarse de una metodología probabilística diferentes corridas del mismo escenario pueden tener resultados diferentes. Por ejemplo, se realizó una corrida 6 veces diferentes manteniendo todas las variables igual. Se corrió durante 90 segundos y cada corrida llegó al mismo resultado (exactamente el óptimo). Sin embargo, los primeros 10 segundos se desarrollaron de la siguiente manera:

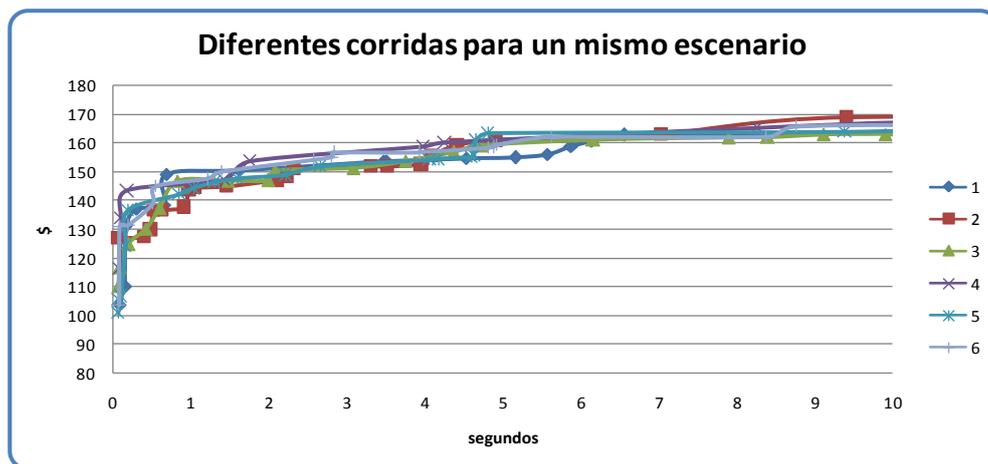


Gráfico 13: Ejemplo de una corrida con el modelo ACO

Cada vez que el sistema encuentra un punto más óptimo que el anterior (es decir un resultado que implique un beneficio mayor) se grafica el punto. Si bien los tiempos y los puntos varían para cada corrida se pueden sacar algunas conclusiones que se ven claramente en el siguiente gráfico. La línea azul muestra el tiempo entre cada nuevo óptimo. La otra línea muestra el incremento porcentual entre óptimos.

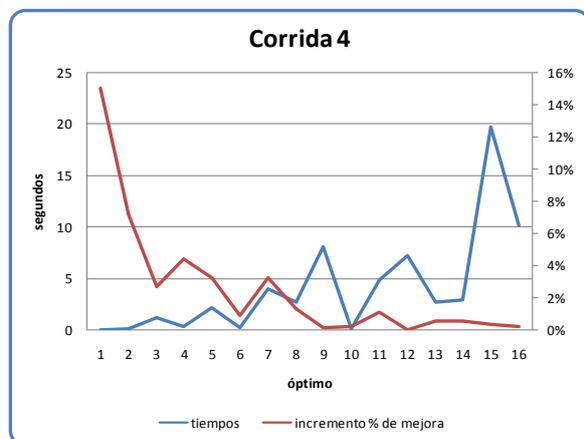


Gráfico 14: Para las nuevas soluciones encontradas: incrementos porcentuales de mejores y tiempos asociados

- El tiempo entre nuevos óptimos aumenta a medida que transcurre la corrida.
- La mejora porcentual entre óptimos disminuye a medida que avanza la corrida.

Esto se cumple en la mayoría de las corridas. Aunque la probabilidad sea baja puede pasar que el primer óptimo de una corrida sea el óptimo global o esté muy cerca del mismo. En ese caso el tiempo hasta encontrar un nuevo óptimo será

muy grande, y la mejora del mismo será muy pequeña. Por otro lado puede pasar que una corrida no llegue a encontrar el resultado óptimo, independientemente del tiempo que se deje corriendo.

Entonces cuanto más tiempo corra el modelo y cuánto más corridas se realicen mayor será la probabilidad de llegar al resultado óptimo. Sin embargo el beneficio de correr el modelo una unidad de tiempo más disminuye a medida que aumenta el tiempo.

Utilizando estas conclusiones y probando diferentes corridas de diferentes magnitudes se llega a las siguientes reglas de decisión:

- A los 2 minutos de la optimización se frena el modelo.
- Si el tiempo desde el último óptimo supera los 20 segundos, se resetea el modelo generando así más de una corrida.
- Como máximo se realizan tres corridas. Si durante la tercera corrida transcurren 20 segundos sin un nuevo óptimo se termina la optimización.

Dado que la mejora porcentual disminuye a medida que avanza el tiempo, si el modelo se frena antes de que el óptimo sea encontrado el resultado será cercano.

2. La validación del output

Una vez determinados todos los parámetros del modelo debe validarse la salida del mismo.

La única manera de saber con certeza cuál es el óptimo de una corrida es corriendo el modelo exhaustivo. Entonces se generaron 100 escenarios aleatoriamente, variando en cada escenario la cantidad de palabras y posiciones y la matriz de valores para ellas. Luego se corrió el modelo de ACO para comparar los resultados.

El rango utilizado para la generación de escenarios de validación fue desde 3 palabras y 3 posiciones hasta 8 palabras y 7 posiciones. En este rango 96 de las 100 veces el modelo ACO llegó al mismo resultado que el modelo exhaustivo. Esto quiere decir que el 96% de las veces el modelo encontró el óptimo²². Además las veces que no llegó al resultado óptimo estuvo cerca (recordemos que en promedio el incremento porcentual de mejora disminuye a medida que se encuentran nuevos óptimos).

²² Daniel Angus en su paper "Solving a unique Shortest PathproblemUsing Ant Colony Optimisation" utiliza un modelo de ACO de un mayor grado de complejidad para resolver un problema de optimización de caminos. Utiliza la misma técnica de validación y confluye que su modelo converge a la solución óptima el 90% de las veces.

Con este experimento queda validada la metodología del modelo de simulación de ACO. Luego se extrapolan las conclusiones al resto de los escenarios usados en el análisis. La razón por la cual no se realiza la validación para un universo más extenso es que el modelo exhaustivo tarda demasiado tiempo. Sólo con 8 palabras y 7 posiciones tarda aproximadamente 9 horas. A continuación se explica el funcionamiento del método exhaustivo de optimización.

4.b Método exhaustivo de optimización

Como el nombre lo explica, este es un método que explora todas las soluciones posibles del sistema. El 100% de las veces se encuentra la solución óptima, sin embargo como el espacio de soluciones se recorre sin ninguna lógica que cuide la eficiencia del proceso, el tiempo implicado puede resultar excesivo.

Se utiliza la misma estructura del modelo de simulación previamente explicado, por ser este el método más rápido entro los métodos probados.

Metodología

1. Se crea la estructura del modelo de simulación.
2. Nacen las hormigas de a una.
3. A cada hormiga se le asigna una combinación y se calcula el resultado de la misma. Nacerán tantas hormigas como combinaciones posibles. Una vez realizado el cálculo muere la hormiga pero si su resultado es mayor al máximo global, se guarda el recorrido.

Tiempo

Para un escenario de I palabras y J posiciones habrán tantas combinaciones como $combinaciones = (J + 1)^I$

A medida que se agrega una palabra o una posición crece el tiempo exponencialmente.

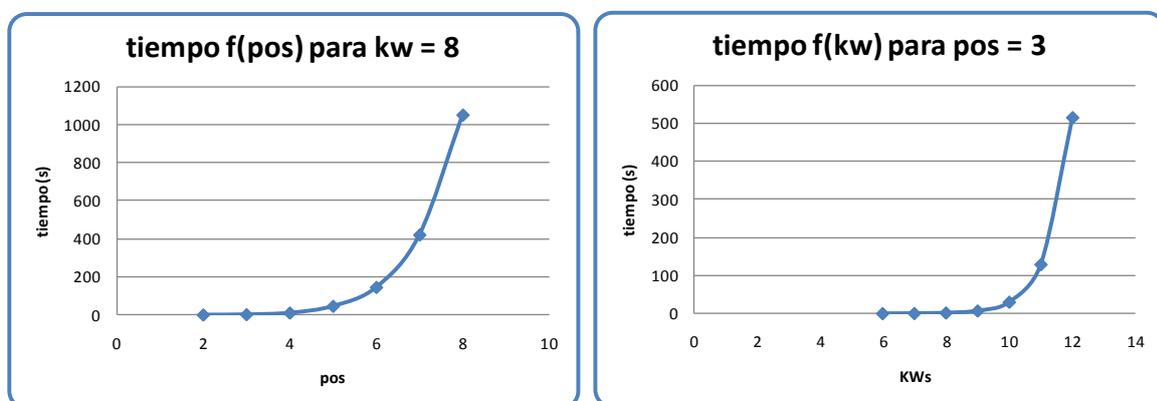


Gráfico 15: tiempo f(pos) & tiempo f(kw)

En el primer gráfico se va que optimizar un escenario de 8 palabras y 8 posiciones toma más de 17 minutos. Con el mismo modelo se optimizó un escenario de 10 palabras y 8 posiciones y tomó aproximadamente 27 horas.

Si $tiempo = T * (J + 1)^I$, se puede encontrar un promedio de T para diferentes combinaciones de J e I para así proyectar cuánto tardaría realizar una optimización exhaustiva de diferentes combinaciones. ¡16 palabras con 5 posiciones tardarían más de 1 año²³! Y 16 palabras es una cantidad totalmente normal para una campaña publicitaria.

Queda claro que el tiempo impide que esta metodología sea utilizada como estrategia. Si bien es la única que permite saber con certeza que el óptimo ha sido encontrado, no cumple los requisitos básicos de las estrategias. Por ejemplo a medida que se aumenta la cantidad de palabras y posiciones, o que se disminuye el horizonte de planeamiento sucede que el tiempo que tarda el modelo en resolverse supera el horizonte temporal de planteamiento.

Es por esto que **la única estrategia que se probará dentro del Tipo 4: Estrategias de Optimización será la estrategia 4a. ACO: optimización basada en colonias de hormigas.**

²³ Este es el tiempo aproximado usando un modelo de simulación como herramienta para realizar el análisis exhaustivo. Se probó el mismo experimento con un código en VBA en Excel que tardaba más de 30% más. Si bien hay otras tecnologías que pueden tardar menos, por ser el crecimiento exponencial y por carecer totalmente de una lógica de búsqueda eficiente, queda claro que los tiempos de este método superan lo aceptable.

5 METODOLOGÍA APLICADA & ANÁLISIS

Este trabajo busca encontrar conclusiones que sean válidas para cualquier escenario, es decir para cualquier industria y para cualquier cantidad de palabras potenciales. Es por eso que se prueba la metodología previamente explicada sobre una variedad de casos tratando de abarcar la mayor cantidad posible de situaciones.

Para cada una de estas situaciones se sigue el siguiente procedimiento. Se muestra con un ejemplo utilizado en el análisis:

- i. Se fija la cantidad de palabras y posiciones
- ii. Se genera la matriz de valores y el presupuesto diario

kw	15
Posiciones	8
Presupuesto	20

INGRESO	posición	0	1	2	3	4	5	6	7	8
keyword										
1		3.4	3.4	3.4	3.4	3.4	3.4	3.4	3.4	3.4
2		30.4	30.4	30.4	30.4	30.4	30.4	30.4	30.4	30.4
3		14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0
4		6.2	6.2	6.2	6.2	6.2	6.2	6.2	6.2	6.2
5		5.6	5.6	5.6	5.6	5.6	5.6	5.6	5.6	5.6
6		9.1	9.1	9.1	9.1	9.1	9.1	9.1	9.1	9.1
7		6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8
8		3.6	3.6	3.6	3.6	3.6	3.6	3.6	3.6	3.6
9		5.5	5.5	5.5	5.5	5.5	5.5	5.5	5.5	5.5
10		5.2	5.2	5.2	5.2	5.2	5.2	5.2	5.2	5.2
11		9.1	9.1	9.1	9.1	9.1	9.1	9.1	9.1	9.1
12		6.1	6.1	6.1	6.1	6.1	6.1	6.1	6.1	6.1
13		6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8
14		14.7	14.7	14.7	14.7	14.7	14.7	14.7	14.7	14.7
15		7.3	7.3	7.3	7.3	7.3	7.3	7.3	7.3	7.3

Costo	posición	0	1	2	3	4	5	6	7	8
keyword										
1		0.0	3.7	2.8	2.3	1.9	1.6	1.4	1.2	1.0
2		0.0	17.7	14.1	12.0	10.6	9.4	8.5	7.7	7.0
3		0.0	13.8	11.2	9.7	8.7	7.8	7.1	6.6	6.1
4		0.0	5.7	4.5	3.9	3.4	3.0	2.7	2.5	2.3
5		0.0	5.4	4.2	3.5	3.0	2.6	2.3	2.0	1.8
6		0.0	8.0	6.6	5.7	5.1	4.6	4.3	3.9	3.6
7		0.0	4.8	3.9	3.4	3.0	2.7	2.5	2.3	2.1
8		0.0	4.0	3.2	2.8	2.4	2.2	2.0	1.8	1.6
9		0.0	4.6	3.8	3.3	2.9	2.7	2.4	2.3	2.1
10		0.0	4.6	3.6	3.0	2.6	2.3	2.0	1.8	1.6
11		0.0	10.3	7.7	6.1	5.0	4.2	3.5	2.9	2.4
12		0.0	4.5	3.7	3.2	2.8	2.6	2.4	2.2	2.0
13		0.0	4.3	3.5	3.1	2.7	2.5	2.3	2.1	2.0
14		0.0	9.0	7.2	6.1	5.3	4.8	4.3	3.9	3.5
15		0.0	5.1	4.1	3.6	3.2	2.9	2.7	2.5	2.3

Volumen	posición	0	1	2	3	4	5	6	7	8
keyword										
1		0	6	3	2	1	1	0	0	0
2		0	4	2	1	1	0	0	0	0
3		0	4	3	2	1	1	0	0	0
4		0	7	4	3	2	1	1	0	0
5		0	6	3	2	1	1	0	0	0
6		0	4	2	1	0	0	0	0	0
7		0	5	3	2	1	1	0	0	0
8		0	5	2	1	0	0	0	0	0
9		0	2	1	0	0	0	0	0	0
10		0	6	3	1	0	0	0	0	0
11		0	6	4	3	2	1	1	1	0
12		0	4	2	1	0	0	0	0	0
13		0	4	2	1	1	0	0	0	0
14		0	5	2	1	1	0	0	0	0
15		0	4	3	2	1	1	0	0	0

- iii. Se calcula el resultado final según cada una de las diferentes estrategias.

Las estrategias son:

Estrategia	Descripción	Resultado (\$)
Tipo 1: Estrategia Equitativa		
1a	Subasta máxima Equitativa	8.5
Tipo 2: Estrategia de Maximización		
2a	Maximización del beneficio total	9.7
2b	Maximización del beneficio/\$	8.4
2c	Maximización del beneficio/\$ * volumen	11.8
Tipo 3: Estrategia de Ranking		
3a	Ranking del beneficio total	13.4
3b	Ranking del beneficio/\$	8.4
3c	Ranking del beneficio/\$ * volumen	14.7
Tipo 4: Estrategia de Optimización		
4a	Optimización ACO	31.6

Tabla 5: Estrategias y resultados asociados (\$)

- iv. Se guardan los resultados: beneficio total, cantidad de palabras, cantidad de posiciones, presupuesto y costo asociado.
- v. Se vuelve a repetir el paso i.

Se prueban todas las estrategias para una gran variedad de escenarios diferentes. En el anexo en el capítulo 6 se pueden ver cada una de estas corridas con sus resultados. En esta sección se muestran los resultados generales.

5.1 Análisis general de las salidas

Normalmente cuando se utiliza publicidad mediante enlaces patrocinados en la red de búsquedas es porque se tiene un objetivo concreto y tangible como la venta de un bien o un servicio. Es por eso que al hablar de ingresos se mencionan las conversiones y es por eso también que el indicador utilizado para comparar las diferentes estrategias es el **beneficio total de la corrida**.

Existen otros objetivos de publicidad como por ejemplo el reforzamiento de la imagen de una marca. Si se utilizara este tipo de publicidad con ese objetivo el resultado debería medirse de otra manera ya que no se simple cuantificar el efecto

de la imagen de marca en una persona. De estar en esta situación se usaría un indicador como la cantidad de impresiones o de *clicks*.

Como primer paso se calcula para cada corrida cuál es el máximo resultado entre las ocho estrategias. Como no se realiza el análisis exhaustivo para cada escenario no se conoce el máximo absoluto, entonces se asume que el mejor resultado entre todas las estrategias es el máximo para ese escenario. A esta estrategia se le asigna un valor de 100% del óptimo y luego se calcula qué porcentaje de ese valor logró cada una de las estrategias. Para cada estrategia se calcula el porcentaje promedio.

1 ^a	2a	2b	2c	3a	3b	3c	4a
51%	54%	23%	52%	60%	23%	64%	99.9%

Tabla 6 Promedio para cada Estrategia

Además se calcula para cada estrategia el intervalo de confianza del 90%, es decir el rango de valores que abarcan al 90% de los resultados, los cuales se grafican a continuación.

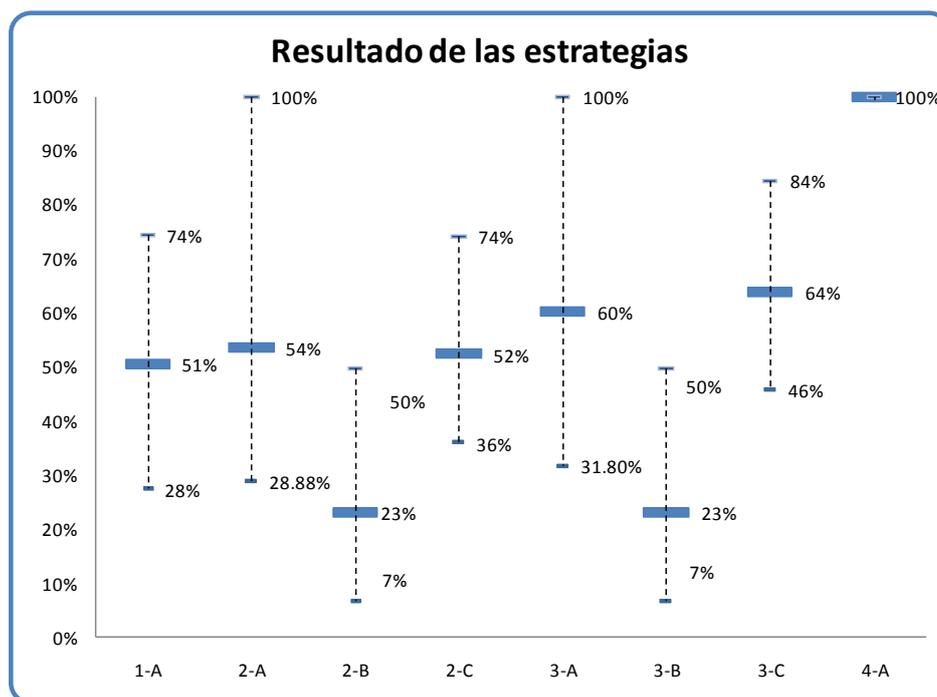


Gráfico 16: Resultado de estrategias

A partir de este gráfico se llega a las siguientes conclusiones:

- La estrategia 4a supera ampliamente al resto de las estrategias, porque tiene un promedio más alto pero además el desvío es prácticamente nulo.
- Las estrategias 2b y 3b tienen prácticamente los mismos valores, y los mismos son muy inferiores al resto de las estrategias.

- La mayoría de las estrategias tienen intervalos de confianza muy amplios. Esto quiere decir que la variabilidad entre los resultados de esa corrida es muy alta. Este puede deberse a que efectivamente el desvío es muy grande o a que dentro de las corridas existen grupos que por sus características tienen resultados muy diferentes.

5.2 Segmentación de las corridas

Se realiza una segmentación para entender si efectivamente existen grupos con resultados diferentes. La variable que se utiliza es la proporción entre el presupuesto de la corrida y la cantidad de palabras.

- i. Si $\frac{\text{presupuesto total}}{\text{cantidad de palabras}} \leq 5 \rightarrow \text{POCO PRESUPUESTO/ kw}$
- ii. Si $\frac{\text{presupuesto total}}{\text{cantidad de palabras}} > 5 \rightarrow \text{MUCHO PRESUPUESTO/kw}$

Debe quedar claro que dos corridas con la misma cantidad de presupuesto diario pueden estar en grupos diferentes: si una tiene muchas palabras puede caer en el grupo de poco presupuesto mientras que si la otra tiene pocas palabras caerá en el en el grupo de mucho presupuesto.

iii. SOBRA PRESUPUESTO

Esto se da cuando se trata de una corrida en la que si eligiéramos la primer posición para cada palabra (que siempre es la más cara) no se llegaría a gastar todo el presupuesto. En otras palabras, en estas corridas el presupuesto no es un factor limitante, es decir que sobra. Es importante resaltar que cuando el presupuesto está mal dimensionado para una campaña, pierde sentido optimizarla.

Entonces los tres grupos que se analizan son:

GRUPOS	
i	POCO PRESUPUESTO/ KW
ii	MUCHO PRESUPUESTO/KW
iii	SOBRA PRESUPUESTO

Los resultados para cada uno de estos grupos son los siguientes:

	1a	2a	2b	2c	3a	3b	3c	4a
POCO	53%	32%	32%	41%	41%	32%	58%	100%
MUCHO	43%	49%	19%	59%	58%	19%	71%	100%
SOBRA	57%	100%	12%	65%	100%	12%	65%	99%

Tabla 7: Resultados segmentados

A partir de estos resultados se realiza el siguiente gráfico radial

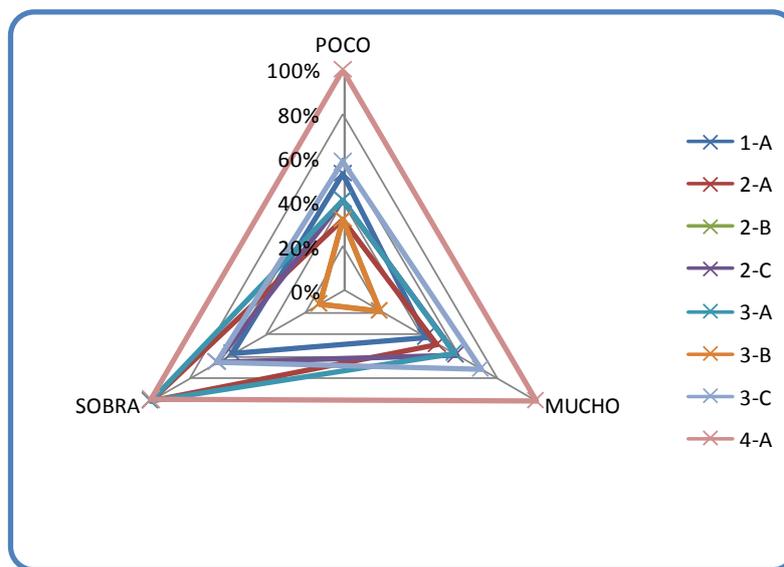


Gráfico 17: Resultado promedio de todas las estrategias en cada grupo

Nuevamente se ve que la estrategia 4a supera (o iguala) al resto en todos los grupos. Sin embargo, se puede ver que las estrategias tienen resultados diferentes en cada grupo, es por eso que se analizan por separado y en más detalle.

5.3 Análisis por Tipo de estrategia

Si bien se ve que la estrategia 4a supera al resto, es importante profundizar también en el resto para entender cuáles sus fortalezas y debilidades

5.3.1 Estrategia Tipo 1: Estrategias cuantitativas

Solamente se probó una estrategia de este tipo: la subasta máxima equitativa. Esta estrategia simplemente divide el presupuesto por igual entre todas las palabras.

La mayor ventaja de esta estrategia es su simplicidad, prácticamente no requiere de esfuerzo de cómputo e incluso podría realizarse sin ningún input de información sobre las palabras (aceptando un empeoramiento de los resultados).

En este caso se le agregó un pequeño grado de complejidad que es que si una palabra implica un resultado negativo en la posición en la que debería ir, la misma se elimina del portfolio de palabras potenciales. Únicamente para esto se requiere la matriz de valores de las palabras.

Sin embargo es una de las estrategias que suelen utilizarse al principio de una campaña, cuando la información sobre las palabras es prácticamente nula.

En promedio esta estrategia llega al 51% del resultado máximo, con una variabilidad muy alta. Al mirar únicamente las corridas del grupo i POCO PRESUPUESTO/KW, el promedio llega al 53% y la variabilidad se reduce un poco.

Como esta estrategia trata a todas las palabras prácticamente por igual, el resultado depende mucho de las características de cada corrida.

Por ejemplo:

- Cuando el presupuesto/ palabra es bajo esta estrategia tiene un rendimiento que no suele ser menor al 40%. Simplemente porque actúa como una estrategia “conservadora”: la oferta de cada palabra será baja y por ende generalmente no saldrán las primeras posiciones. Si bien la posición resultante depende de la curva de costos de cada palabra, en promedio salen las posiciones del medio. Llamamos a este escenario conservador porque indirectamente cuida la eficiencia de la inversión (recordando que la misma aumenta a medida que aumentamos de posición).
- Cuando el presupuesto/palabra es alto suelen salir las primeras posiciones. En este caso el resultado depende directamente del rendimiento de estas palabras que siempre tienen un buen volumen, pero que pueden tener malos márgenes.
- Cuanto más similares son las matrices de valores de las palabras, más extremos son los resultados (o muy buenos, o muy malos).

5.3.2 Estrategia Tipo 2: Estrategias de maximización por palabra

Este tipo de estrategias tiene un grado mayor de lógica que la anterior: para cada palabra se busca optimizar algún indicador.

Los indicadores utilizados con sus respectivas lógicas fueron:

- a. Beneficio total: mira el resultado final de cada palabra, no tiene en cuenta la eficiencia con la que se invierte cada unidad de dinero.
- b. Beneficio/\$: mira únicamente la eficiencia con la que se invierte el dinero.
- c. Beneficio/\$ * volumen: intenta combinar ambos objetivos.

Los resultados fueron los siguientes

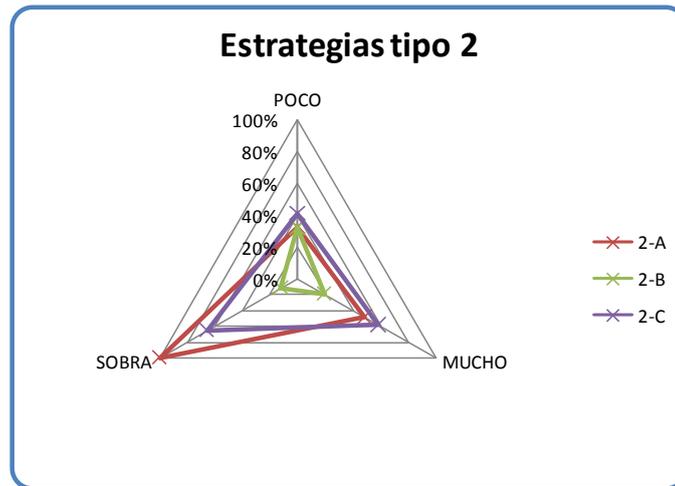


Gráfico 18: Comparación de grupos para la estrategia tipo 2

La estrategia 2B puede ser descartada desde el principio. Sucede que la posición que maximiza este indicador es siempre la última, y el volumen para esta posición siempre es muy pequeño. Entonces si bien se invierte el dinero eficientemente, nunca se llega a usar todo el presupuesto y el resultado es siempre mucho menor al óptimo.

Comparando las estrategias 2A y 2C se identifican dos casos diferentes: el primero conformado por el grupo en el que el presupuesto sobra, el segundo por los otros dos.

En el primer caso la estrategia 2A supera a la 2C. Este caso ocurre cuando sobra presupuesto. En estas condiciones la combinación óptima ocurre cuando maximizamos el resultado total de cada palabra y esto tiene sentido porque como el presupuesto no es un factor limitante no hay necesidad de optimizar la eficiencia con la que se invierte.

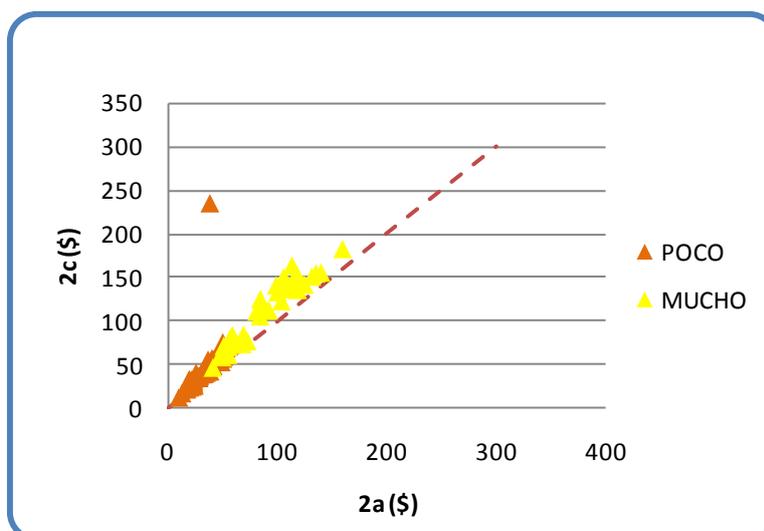


Gráfico 19: Comparación de corridas 2A vs 2C

El segundo es el caso de mayor interés. Cuando el presupuesto es un factor limitante y debemos optimizar su uso, sí es interesante comparar ambas estrategias. A la izquierda están graficadas todas las corridas ya no en valor porcentual sino que en valor absoluto. Se ve que cuando el presupuesto sí es un factor limitante conviene la estrategia 2C antes que la 2A. Esto quiere decir que conviene optimizar con el beneficio total/ \$ invertido antes que con el beneficio total.

5.3.3 Estrategia Tipo 3: Estrategias de ranking

Este tipo de estrategias tiene un grado todavía mayor de integración del sistema. Se vuelven a probar los mismos tres indicadores:

- Beneficio total
- Beneficio/\$.
- Beneficio/\$ * volumen

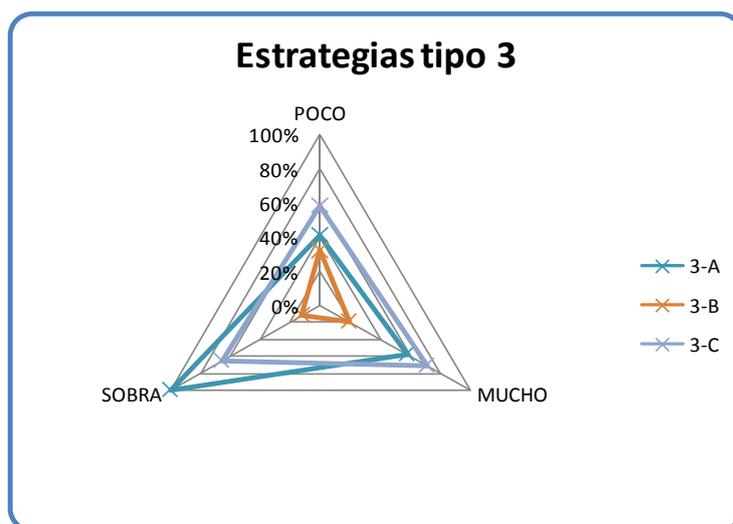


Gráfico 20: Comparación de Estrategias tipo 3

entre usar el indicador A y el indicador C cuando el presupuesto no sobra es todavía mayor.

Nuevamente conviene utilizar el indicador C (beneficio total/\$) siempre que el presupuesto no sobra. Comparamos entonces la estrategia 2C contra la 3C.

Los resultados son parecidos al caso anterior.

- Cuando el presupuesto sobra los resultados son exactamente igual que en el caso anterior: como el presupuesto nunca se agota, no se descartan palabras del ranking.
- Por la misma razón la estrategia 3B es exactamente igual a la estrategia 2B y nuevamente se descarta.

- Esta vez la diferencia

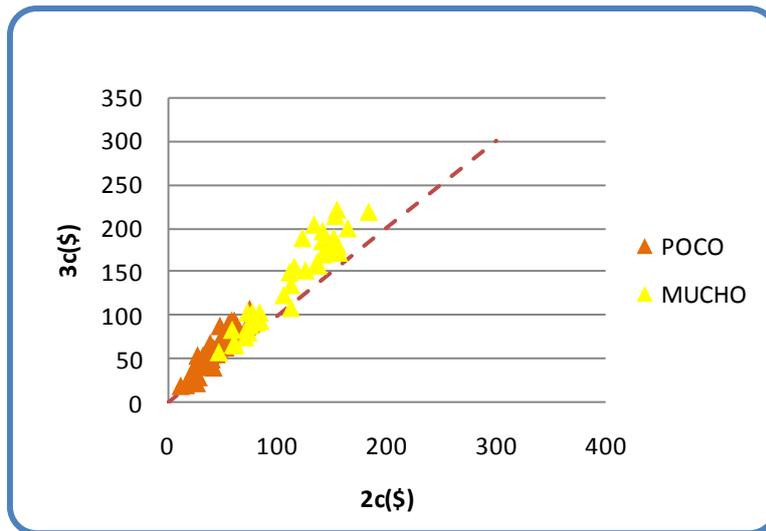


Gráfico 21: Comparación de corridas 2C vs 3C

Tal como era de esperarse la estrategia 3C supera la 2C. Es lógico ya que como ya fue explicado las estrategias del tipo 3 tienen un mayor grado de integración del sistema. En vez de optimizar cada palabra se mira todas las palabras en conjunto y únicamente se agrega una nueva palabra si el resultado total del sistema mejora. Esta metodología permite la exclusión de palabras mientras que la otra no.

Juntando las estrategias de tipo 1, 2 y 3 se obtiene el siguiente gráfico:

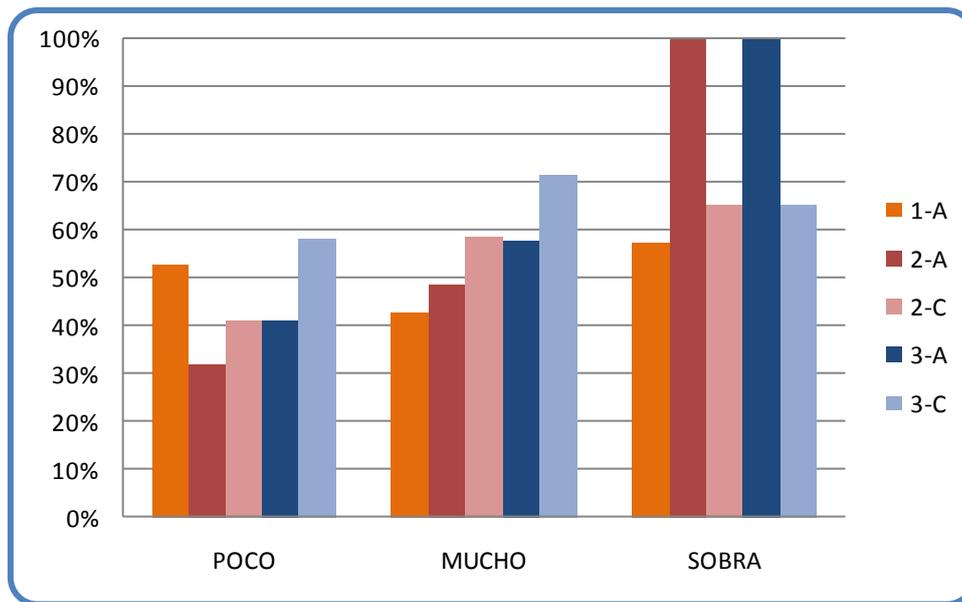


Gráfico 22: Resumen de las estrategias tipo 1, 2 y 3

Las estrategias 2B y 3B han sido eliminadas por su performance extremadamente pobres. Las conclusiones principales para las estrategias tipo 1, 2 y 3 son las siguientes:

- Cuando el presupuesto sobra conviene utilizar el indicador beneficio total. Las estrategias 2A y 3A dan el mismo resultado que de hecho es el óptimo del sistema. Sin embargo la optimización en este caso carece de importancia ya que el presupuesto no es un factor limitante.
- En el resto de los casos el indicador a usar para optimizar es el beneficio total/costo. La estrategia que conviene utilizar es la de tipo 3 (realizando un ranking y agregando de a una palabra siempre que esta mejore el resultado final).
- En el primer grupo la estrategia 1A da un resultado muy cercano al mejor los graficados.

Eliminando las corridas en las cuales sobra el presupuesto se obtienen los siguientes intervalos de confianza. En la mayoría de los casos los intervalos han sido reducidos con respecto al caso anterior que incluía todas las corridas. Sin embargo los mismos siguen siendo amplios.

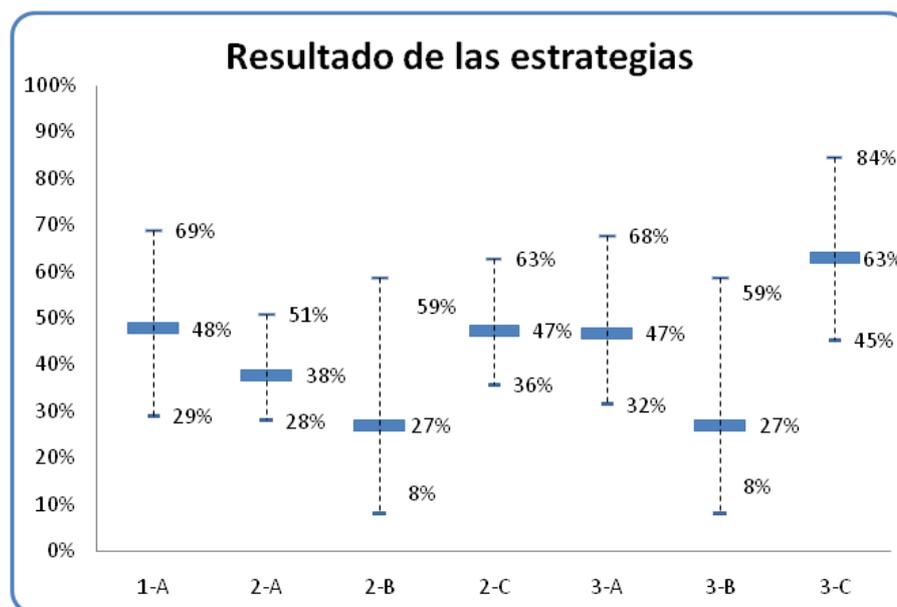


Gráfico 23: Estrategias tipo 1, 2 y 3 descartando las corridas en las cuales sobra presupuesto

Hasta ahora, descartando las corridas en las cuales sobra el presupuesto, ninguna estrategia llega a un promedio del 65%. ¿Cómo es esto posible? ¿Por qué no se puede llegar a mejores resultados con las estrategias propuestas?

Hay dos razones principales:

- i. No se encuentra un indicador único que sirva de optimizador. Esto es porque hay que cuidar tres factores a la vez: el ingreso, el costo y el volumen. Y además hay un presupuesto que hay que optimizar.
- ii. Las metodologías propuestas corren el riesgo de caer en óptimos locales en vez de globales. De los tres tipos de estrategias analizados hasta ahora el tipo 3 es el más complejo. Este método permite al sistema descartar palabras que no mejoren el resultado. Sin embargo, no permite al sistema elegir un sub-óptimo para una palabra en pos de encontrar un óptimo global.

5.3.4 Estrategia tipo 4: optimización global

Lógicamente la estrategia tipo 4 sobrepasa los problemas recién expuestos. ¡De eso se tratan las optimizaciones globales!

Cuando el presupuesto por palabra potencial es poco o mucho, esta estrategia encuentra el mejor resultado el 100% de las corridas. Esto quiere decir que el promedio es del 100% y el desvío es nulo.

En cambio, cuando el presupuesto sobra esta estrategia no siempre llega al óptimo. De todas las corridas de este grupo, un 20% de las veces la optimización no encontró el óptimo. Sin embargo en promedio llegó al 98% del mismo²⁴.

5.4 *Análisis más allá del resultado*

Si bien en todo momento se comparó las estrategias según el resultado final, hay otros factores que influyen en la eficiencia de una estrategia frente a otra.

5.4.1 **Tiempo**

El tiempo siempre es un factor a tener en cuenta, porque se traduce directamente en costo.

- Las estrategias tipo 1, 2 y 3 se realizan instantáneamente.
- La estrategia tipo 4 tarda como máximo 2 minutos y en promedio las corridas duran 90 segundos. Si bien no es instantáneo se considera un tiempo despreciable frente a la duración del horizonte temporal de planeamiento y del beneficio que se obtiene.

5.4.2 **Complejidad**

- La estrategia tipo 1 tiene un proceso iterativo pero el poder de cálculo necesario para realizarlo es mínimo. Incluso se podría usar esta estrategia sin información previa sobre las palabras y únicamente en el caso de obtener un resultado negativo se eliminaría la palabra (sin embargo esto implica un riesgo de un posible costo asociado a usar una palabra en una posición que tiene rendimiento negativo).
- La estrategia tipo 2 requiere el cálculo de un indicador para cada posición de cada palabra. Luego simplemente se calcula el máximo para cada palabra.
- La estrategia tipo 3 además utiliza un ranking y un proceso iterativo para ver hasta qué palabra se agrega
- La estrategia tipo 4 es sin lugar a dudas la más compleja. Por un lado requiere conocimiento avanzado para programar el modelo de simulación. Además el modelo necesita un poder de cálculo mayor que el resto de las estrategias para ser ejecutado. Si bien prácticamente cualquier computadora actual tiene la potencia requerida se necesita un software específico para ejecutar el modelo de simulación.

²⁴La razón por la que esto sucede es que este modelo fue optimizado para escenarios en los que el presupuesto era un factor limitante. En la práctica para que el modelo funcione en estos casos hay que darle un peso más chico a la feromona.

5.5 Resumen del análisis

	Resultado	Tiempo	Complejidad
TIPO 1: Subasta máxima equitativa	Sus resultados nunca son los mejores. Sin embargo son razonablemente buenas cuando el presupuesto/ palabras potenciales es bajo (en promedio 53%).	Instantáneo	Simple. Únicamente utiliza un proceso iterativo.
TIPO 2: Maximización por palabra	Cuando el presupuesto sobra, esta estrategia sirve para encontrar el óptimo (promedio = 100%). El indicador a utilizar es el beneficio total	Instantáneo	Simple. Tiene que calcular un indicador para cada posición de cada palabra y encontrar el máximo para cada palabra.
TIPO 3: Ranking	Cuando el presupuesto no sobra, esta estrategia arroja un resultado razonable (promedio = 63%). Si bien no es el óptimo es la mejor estrategia después de la optimización global. El indicador a utilizar es el beneficio total/costo	Instantáneo	Simple. Tiene que calcular un indicador para cada posición de cada palabra, realizar un ranking, y realizar un proceso iterativo.
TIPO 4: Optimización global	El promedio total es de 99.9%. Además es en promedio un <ul style="list-style-type: none"> • 145% mayor a la estrategia tipo 1 • 104% mayor a la estrategia 2C • 63 % a la estrategia 3C. 	Como máximo 2 minutos.	Complejo. Requiere un conocimiento alto en el lenguaje de simulación para la programación, pero una vez que está programado es simple de ejecutar. Requiere un software específico.

Tabla 8: Resumen del análisis de la metodología aplicada

6 CONCLUSIONES Y FUTURAS LÍNEAS DE INVESTIGACIÓN

6.1 Conclusiones de la tesis

El objetivo principal de esta tesis era poder identificar, explicar y analizar diferentes estrategias de manejo de presupuesto en la publicidad mediante enlaces patrocinados en la red de búsqueda.

Las estrategias debían ser: automáticas, consistentes, extrapolables y reproducibles.

Se acotó el análisis a cuatro tipos de estrategias que fueron:

- Tipo 1: Estrategia Equitativa
- Tipo 2: Estrategia de Maximización
- Tipo 3: Estrategia de Ranking
- Tipo 4: Estrategia de optimización

A lo largo del proyecto surgieron diferentes deducciones que vale la pena resaltar.

En primer lugar se habló del índice de calidad de los anuncios y se concluyó que es indispensable llevar este índice tan alto como se esté dispuesto (en términos de dinero y tiempo) antes de intensificar el esfuerzo puesto en optimizar las subastas. Esto es porque como ya fue explicado el beneficio que resulta de mejorar el índice es sustancialmente mayor que el beneficio que se puede sacar de realizar las subastas de manera más óptima.

También se explicó que para este análisis el indicador utilizado para comparar estrategias es el beneficio total de una combinación donde

$$\text{beneficio total} = (\text{ingreso}_{\text{por click}} - \text{costo}_{\text{por click}}) \times \text{clicks potenciales}$$

y de ser necesario el mismo se ajusta al presupuesto diario. Entonces este análisis está enfocado a casos en que se vende un producto o un servicio con un ingreso asociado.

Se resaltó además la importancia de tener campañas bien dimensionadas. Esto quiere decir que debe haber un buen balance entre la cantidad de palabras claves y el presupuesto destinado a la campaña. El valor depende mucho de la industria en la que se esté trabajando y del tipo de palabras o combinaciones de palabras que se use, pero la conclusión es que si hay presupuesto ocioso no sólo no se podrá gastar todo sino que lo que se gastará se hará de forma no eficiente.

Lograr este balance es un proceso iterativo. Cuando se comienza una campaña es muy poca la noción que se tiene sobre las matrices de valores de las palabras y sobre cuál sería un presupuesto adecuado. Es cuestión de no olvidarse que cuando el presupuesto sobra se podría invertir de manera más eficiente en otros tipos de publicidad.

Brendan Kitts y Benjamin Leblanc²⁵ comparan el resultado de una campaña manejada por un experto y la misma campaña manejada usando subastas óptimas. Concluyen que el resultado en el segundo caso es del orden de cuatro veces mayor que en el primero.

En esta tesis se utiliza un enfoque diferente. En lugar de usar el resultado de un experto como resultado base se utilizan una serie de estrategias. La optimización global (estrategia tipo 4) equivale a la estrategia propuesta por los autores. Esta obtiene resultados en promedio 4 veces mayores al resto de las estrategias.

Cuando el presupuesto es un factor limitante esta estrategia llega al mejor resultado el 100% de las veces. Cuando no es un factor limitante encuentra el mejor resultado el 80% de las veces, pero el 20% restante llega en promedio al 98% del mejor valor.

La estrategia que le sigue en cuanto a resultados es la estrategia 3C (estrategia de ranking usando el indicador beneficio unitario por \$). La optimización global obtiene resultados en promedio un 65% mayores.

No caben dudas que la optimización global obtiene resultados superiores al resto, sin embargo hay otros factores a analizar como complejidad y tiempo. Desde esta perspectiva se ve que la complejidad de la optimización global puede llegar a ser un factor limitante para su aplicación.

La conclusión subyacente de este análisis es que vale la pena abordar el tema de las subastas en Adwords desde el enfoque planteado; es decir entendiendo al sistema completo y cómo funciona, y entendiendo además que existe un óptimo y que es posible encontrarlo o acercarse a él. El grado de complejidad con que se hace no será siempre el mismo pero el enfoque del problema sí lo será

6.2 Conclusiones generales de la metodología utilizada

Si bien el análisis se realizó para un caso puntual que se refiere a los enlaces patrocinados en la red de búsquedas con el objetivo de la venta de un bien o servicio; el valor de los resultados puede extrapolarse a un caso más general. De

²⁵Brendan Kitts, Benjamin Leblanc. *Optimal Bidding on Keyword Auctions*. *Electronic Markets*, 14(3):186-201, September 2004.

alguna manera lo que se estudió fue la aplicación de una metodología a un caso puntual. Vale la pena entonces resaltar las conclusiones de la metodología en sí.

La estrategia Tipo 1, estrategia equitativa, no contempla el concepto de optimización pero sí busca un equilibrio entre todas las palabras. En este caso se equilibró la subasta máxima entre todas las palabras lo cual resultó ser una estrategia muy simple que prácticamente no requería de información. Siempre que se comparan estrategias es conveniente tener una opción más simple que el resto ya que la complejidad puede ser un factor limitante. En este análisis, para un caso particular, se obtuvo un resultado contra-intuitivo ya que esta opción resultó mejor que algunas de las otras que de hecho eran más complejas.

La estrategia Tipo 2, estrategia de maximización, optimiza cada palabra por separado. Incorpora el concepto de subastar utilizando una optimización, pero al hacerlo para cada palabra por separado no lo hace de la mejor manera.

La estrategia Tipo 3, estrategia de ranking, ya aumenta el grado de integración del sistema porque luego de realizar un ranking de algún indicador agrega palabras sí y sólo sí el resultado del sistema mejora. Por dos razones este enfoque no es suficiente para llegar al mejor resultado. Por un lado, no siempre existe un solo indicador que por sí solo contemple todas las relaciones del sistema. Por ejemplo en el caso de análisis de esta tesis habían dos objetivos contrapuestos: eficiencia de inversión (contemplada mediante el beneficio por \$ invertido) y volumen (contemplado mediante *clicks* o beneficio total). Realizar el ranking usando más de un indicador es complejo y requiere de otros parámetros. La segunda razón es que a veces hay que situar a una palabra en un sub-óptimo en pos de encontrar el óptimo global del sistema. Esto se traduce en que a veces convendría saltarse una palabra del ranking para obtener un mejor resultado final y esta opción no es contemplada en la metodología.

La estrategia Tipo 4, estrategia de optimización global, integra el sistema en su totalidad. Si cambiaran los objetivos de la publicidad cambiaría el indicador utilizado y posiblemente sería necesario volver a optimizar los parámetros pero como el modelo fue validado se puede tener certeza que volvería a funcionar.

6.3 Limitaciones del estudio

La limitación principal de este estudio es el gran grado de simplificación que se utilizó. Entre otras cosas:

- Se utilizó un horizonte temporal de planeamiento diario, lo cual facilita la definición de las estrategias por tratarse de un servicio que utiliza un presupuesto diario.
- Las curvas utilizadas se consideraron constantes en el tiempo.

- Las curvas teóricas utilizadas no tuvieron ningún ruido, fueron funciones exponenciales perfectas.
- El índice de calidad fue constante.

La razón por la cual se realizaron estas simplificaciones es que se buscaba un sistema teórico lo suficientemente representativo de la realidad como para poder comparar las estrategias.

Los resultados del análisis con estas simplificaciones son válidos en términos relativos. Es decir, el porcentaje de mejora de una estrategia frente a otra es válido. Ahora bien podrían desarrollarse algunos de estos puntos para lograr resultados todavía más representativos de la realidad. Y con esta idea nacen las futuras líneas de investigación a continuación.

6.4 Futuras líneas de investigación

Las futuras líneas de investigación se encuentran en dos niveles diferentes. Por un lado y como fue explicado en la sección anterior, están centradas en complejizar el sistema teórico para acercarlo a la realidad. Por el otro tratan sobre expandir el alcance del análisis.

En cuanto a complejizar el sistema teórico, el mayor potencial se encuentra en acortar el horizonte temporal de planeamiento. Esto en la práctica significa tomar decisiones más de una vez por día. Para realizar esto sería necesario estimar las curvas en diferentes momentos del día; luego el modelo de simulación sería extrapolado a esta situación agregándoles nodos. En vez de la situación actual que tiene un nodo por cada palabra-posición, habría un nodo por cada palabra-posición-tiempo (llamando tiempo a cada punto en que se toma una decisión).

La expansión del alcance del análisis se refiere a la obtención de la matriz de valores de las palabras. Esto es necesario para desarrollar la implementación de las estrategias propuestas. La base de la obtención de la matriz de valores es clara: se relevan diferentes puntos y luego se encuentran los parámetros de la función exponencial que minimizan el error mediante mínimos cuadrados. Esto se aplica tanto a la función de costo y a la función de volumen. Sin embargo hay diferentes cuestiones que podrían ser investigadas:

- Cada cuánto cambian las curvas: Google Adwords proporciona a sus usuarios indicadores diarios de sus campañas (precio promedio, cantidad de *clicks*, posición promedio, etc.) de modo que puede obtenerse un punto de la curva por día. Ahora bien, las curvas cambian en el tiempo porque cambia la cantidad de competidores, sus índices de calidad y los precios que éstos ofertan. Por esta razón es importante entender cuántos puntos son necesarios para obtener

curvas confiables, cada cuánto conviene descartar puntos y cuál es la manera más eficiente de obtenerlos.

- Cómo sacarle provecho a los nuevos puntos: una vez que se usa una optimización y se implementa la misma se obtienen nuevos puntos. Cada nuevo punto puede utilizarse de dos maneras. Por un lado pueden volverse a ajustar los parámetros de las curvas agregando el nuevo punto. Además puede compararse el resultado real con el resultado arrojado por la estrategia para ajustar otros parámetros haciendo el proceso iterativo. El desarrollo de esta línea del proceso es fundamental y de gran potencial.

La última línea de investigación está orientada a complementar los enlaces patrocinados en la red de búsqueda con otros tipos de publicidades online. A saber, enlaces patrocinados en la red de display, banners, y publicidad en redes sociales (como por ejemplo Facebook).

7 BIBLIOGRAFÍA

Introducción

- Internet World Stats: <<http://www.internetworldstats.com/stats.htm>>Página vigente al 05/03/2011.

Marco teórico

- Richardson George, Pugh, Alexandre L. 1981. **Introduction to System Dynamics Modeling**, Productivity Press. ISBN: 0-915299-24-0
- Shannon, Robert; Johannes, James D. (1976), **Systems simulation: the art and science**. IEEE Transactions on Systems, Man and Cybernetics
- Historia de Google: <http://publishing2.com/2008/05/27/google-adwords-a-brief-history-of-online-advertising-innovation/>Página vigente al 05/03/2011.
- Nivel de calidad: <<http://adwords.google.com>>
- Frederick S. Hillier y Gerald J. Lieberman. 2010. **Investigación de Operaciones**. Editorial Mc Graw Hill. ISBN: 970-10-3486-4.
- Winston, W. 2004. **Investigación de Operaciones**. Editorial Cengage. ISBN: 9706863621.
- Universidad Internacional de Andalucía, **Indorucción a los algoritmos meta-heurísticos**. Granada, España 2006. <<http://sci2s.ugr.es/seminars/5/Int-Metaheurísticas-UIA-Baeza-2006.pdf>>
- Dorigo M., V. Maniezzo& A. Colorni.**Positive Feedback as a Search Strategy**. *TechnicalReport No. 91-016*, Politecnico di Milano, Italy. 1991
- Dorigo, M, **Optimization, Learning and Natural Algorithms**, PhD thesis, Politecnico di Milano, Italy, 1992.
- Daniel Angus, **Solving a unique Shortest Path problem using Ant Colony Optimisation**
- Daniel C. Asmar, Ahmad Elshamli, ShawkiAreibi. **A Comparative Assessment of ACO Algorithms with a TSP Environment**. 2007.
- David Segal, **The Dirty Little Secrets of Search**, The New York Times, 2011-02-12.
- LudekCygler. **Semester Project: Bidding Agent for Advertisement Auctions**, 2008-2009.

Contexto

- Evolución del CPC: <www.optimum7.com/internet-marketing/search-engine-marketing/average-adwords-pay-per-click-ppc-costs.html>Página vigente al día 10/07/2011.
- *Internet Advertising Bureau*: <<http://www.iab.net>>Página vigente al día 10/07/2011.

Metodología

- Brendan Kitts, Benjamin Leblanc, **Optimal Bidding on Keyword Auctions**. *ElectronicMarkets*, 14(3):186-201, September 2004.
- Borgs, Chayes, Etesami, Immorlica, Jain, Mahdian, **Bid optimization in online advertisement auctions**. Microsoft Research.
- Xj Technologies, Simulation Software and Services, <www.xjtek.com/anylogic>

8 ANEXO

8.1 Resultados completos

A continuación se explicitan los resultados crudos de todas las corridas. Cada corrida tiene su propia matriz de valores para cada palabra y en la tabla se ve el resultado para cada estrategia. A la derecha se ve la cantidad de palabras, la cantidad de posiciones y el presupuesto diario en cada caso.

Corrida/ Estrategia	1a	2a	2b	2c	3a	3b	3c	4a	kw	pos	presu
1	5.0	17.5	4.8	12.7	17.5	4.8	12.7	17.5	5	8	120
2	62.8	198.6	42.3	113.6	198.6	42.3	113.6	198.6	20	8	480
3	31.5	171.5	30.0	81.7	171.5	30.0	81.7	171.5	20	8	800
4	276.1	406.1	35.2	268.6	406.1	35.2	268.6	406.1	25	8	1000
5	276.1	406.1	35.2	268.6	406.1	35.2	268.6	404.9	25	8	8000
6	29.7	109.9	18.4	66.7	109.9	18.4	66.7	109.9	16	8	768
7	88.6	178.7	29.5	129.9	178.7	29.5	129.9	178.7	16	8	768
8	123.5	223.0	24.5	160.2	223.0	24.5	160.2	223.0	16	8	1000
9	95.9	180.0	21.2	120.7	180.0	21.2	120.7	180.0	16	8	1000
10	104.7	198.1	31.5	143.3	198.1	31.5	143.3	198.1	16	8	768
11	126.9	228.1	43.7	148.6	228.1	43.7	148.6	228.0	16	8	768
12	153.1	272.9	2.1	197.6	272.9	2.1	197.6	272.9	16	20	384
13	8.2	89.4	1.1	19.3	89.4	1.1	19.3	86.5	16	20	384
14	105.3	257.6	34.4	158.5	257.6	34.4	158.5	258.8	20	8	480
15	1464.1	1646.2	109.2	1272.3	1646.2	109.2	1272.3	1646.2	50	8	2400
16	1140.8	1454.1	121.9	938.5	1454.1	121.9	938.5	1406.2	50	8	2400
17	531.7	735.3	54.1	528.9	735.3	54.1	528.9	720.4	40	8	1920
18	474.7	763.2	63.6	507.5	763.2	63.6	507.5	825.8	40	8	960
19	480.5	611.8	37.1	485.8	611.8	37.1	485.8	611.8	25	8	1200
20	158.9	327.6	38.0	208.2	327.6	38.0	208.2	327.6	25	8	500
21	345.8	500.6	40.4	368.0	500.6	40.4	368.0	500.6	25	8	600
22	20.8	51.8	12.4	33.6	51.8	12.4	33.6	51.8	10	8	240
23	101.3	158.6	24.1	124.0	158.6	24.1	124.0	158.6	15	8	360
24	1.3	43.2	10.7	19.8	43.2	10.7	19.8	43.2	8	8	192
25	11.4	43.2	10.7	19.8	43.2	10.7	19.8	43.2	8	8	50
26	2685.9	3013.7	608.2	1898.8	3013.7	608.2	1898.8	2865.0	60	5	2880
27	371.3	561.2	125.9	342.2	561.2	125.9	342.2	556.9	30	6	1440
28	186.8	295.4	67.8	185.0	295.4	67.8	185.0	295.4	20	6	500
29	482.0	648.9	56.8	478.4	648.9	56.8	478.4	643.7	25	8	1000
30	31.2	125.9	39.5	134.5	138.9	39.5	158.1	218.8	20	8	160
31	49.2	97.5	14.3	78.8	85.2	14.3	69.1	152.6	20	8	160
32	275.2	355.1	47.2	376.7	365.4	47.2	381.3	432.9	20	8	480
33	33.5	88.6	43.9	94.2	92.3	43.9	100.1	242.1	20	8	160
34	93.3	27.7	43.9	29.4	28.9	43.9	31.3	137.4	20	8	50
35	149.8	261.1	32.1	189.3	261.1	32.1	189.3	261.1	20	8	480
36	55.9	139.3	42.2	96.4	167.8	42.2	96.4	216.6	20	8	160
37	196.8	295.5	21.0	229.7	295.5	21.0	229.7	295.5	20	8	480
38	80.2	65.0	21.0	76.2	109.5	21.0	97.3	166.0	20	8	100
39	240.1	350.8	29.4	249.5	350.8	29.4	249.5	350.8	20	8	480
40	56.6	69.8	29.1	74.7	73.1	29.1	92.9	145.0	20	8	100
41	135.5	115.1	43.2	115.8	123.0	43.2	133.0	278.4	25	8	200
42	121.4	28.8	43.2	28.9	30.8	43.2	33.3	146.5	25	8	50
43	102.9	69.1	43.2	69.5	73.8	43.2	79.8	221.1	25	8	120

44	228.0	345.3	30.2	318.4	341.8	30.2	318.4	398.5	25	8	600
45	33.0	186.8	47.0	200.1	214.5	47.0	246.5	362.9	25	8	200
46	132.0	122.3	41.2	132.5	176.9	41.2	189.9	287.0	25	8	200
47	83.0	30.6	41.2	33.1	38.2	41.2	67.9	135.3	25	8	50
48	243.3	464.5	60.7	376.2	474.0	60.7	376.2	518.8	25	8	500
49	79.1	143.8	53.5	150.3	150.4	53.5	180.8	290.2	25	8	200
50	78.3	111.0	58.6	127.2	124.4	58.6	142.2	279.7	25	8	200
51	298.3	348.4	33.5	328.9	360.6	33.5	328.2	476.0	25	8	500
52	365.3	503.4	33.5	371.4	503.4	33.5	371.4	499.8	25	8	1000
53	136.9	69.7	33.5	65.8	80.8	33.5	51.8	213.3	25	8	100
54	59.9	142.0	31.3	147.7	148.2	31.3	157.2	259.7	25	8	200
55	169.3	355.0	33.8	209.8	355.0	33.8	209.8	355.0	26	8	624
56	227.6	352.7	53.1	308.2	361.2	53.1	308.2	425.7	26	8	500
57	62.5	156.1	42.0	148.3	206.8	42.0	205.3	333.2	26	8	208
58	243.9	352.9	28.3	290.8	359.2	28.3	290.8	388.8	26	8	500
59	26.6	147.9	23.0	130.2	165.8	23.0	130.2	220.7	26	8	208
60	47.5	114.8	29.9	105.5	107.3	29.9	100.4	219.6	26	8	208
61	102.2	307.0	34.7	169.1	307.0	34.7	169.1	307.0	26	8	500
62	161.1	324.1	31.2	169.0	324.1	31.2	169.0	323.6	26	8	624
63	36.0	193.0	31.2	169.0	196.9	31.2	169.0	282.5	26	8	200
64	180.1	349.2	32.1	245.0	349.2	32.1	245.0	349.2	26	8	500
65	127.3	71.0	32.1	69.3	75.4	32.1	80.4	200.7	26	8	100
66	146.9	196.5	31.7	205.2	218.1	31.7	211.0	294.6	26	8	208
67	294.8	464.9	50.3	314.6	464.9	50.3	314.6	463.5	29	8	800
68	216.9	341.6	44.5	244.8	341.6	44.5	244.8	339.9	29	8	900
69	209.7	321.7	44.5	244.8	311.9	44.5	244.8	335.4	29	8	600
70	9.5	160.8	44.5	149.5	174.0	44.5	166.6	281.9	29	8	300
71	109.4	26.8	44.5	24.9	30.7	44.5	45.9	145.8	29	8	50
72	189.0	80.5	41.0	62.3	117.2	41.0	101.6	274.4	29	8	150
73	108.5	182.6	42.2	182.4	169.8	42.2	176.4	328.6	29	8	232
74	181.8	328.2	58.7	377.3	366.9	58.7	396.2	537.2	29	8	500
75	80.8	133.3	60.3	99.5	145.1	60.3	155.3	343.7	29	8	232
76	0.0	135.8	28.3	114.4	118.6	28.3	114.4	195.7	20	8	160
77	75.1	41.6	37.4	42.7	51.8	37.4	56.1	140.1	20	8	50
78	46.2	81.6	23.4	70.7	102.6	23.4	102.8	202.3	20	8	200
79	865.6	1370.4	137.2	1000.8	1455.8	137.2	1000.8	1597.9	60	8	1440
80	538.9	114.2	137.2	128.0	157.4	137.2	249.6	582.5	60	8	120
81	274.6	329.7	68.2	369.3	324.5	68.2	329.2	799.7	50	8	400
82	170.8	382.8	51.8	416.2	409.8	51.8	416.2	616.8	40	8	500
83	347.3	694.1	69.0	525.4	666.9	69.0	525.4	753.5	40	8	960
84	260.1	72.3	69.0	71.7	63.3	69.0	123.5	311.8	40	8	100
85	168.5	431.0	105.5	431.0	420.4	105.5	431.0	519.0	40	8	500
86	236.3	396.8	60.1	419.7	327.8	60.1	419.7	613.9	40	8	600
87	262.7	132.3	60.1	144.1	113.7	60.1	200.6	439.2	40	8	200
88	303.9	703.1	82.6	440.6	701.4	82.6	440.6	780.6	40	8	800
89	137.1	419.5	94.0	458.2	444.3	94.0	533.7	735.7	40	8	500
90	303.5	83.9	94.0	91.6	87.2	94.0	184.2	368.8	40	8	100
91	272.1	393.8	81.6	440.8	388.0	81.6	511.5	776.4	40	8	600
92	182.0	240.5	61.0	238.7	304.2	61.0	334.9	584.9	40	8	320
93	197.9	305.6	94.2	360.9	271.9	94.2	443.6	648.6	40	8	320
94	281.8	450.1	118.1	475.3	461.1	118.1	555.2	887.1	40	8	500
95	282.3	234.8	80.0	190.4	220.1	80.0	247.2	608.7	40	8	320
96	219.3	483.9	86.3	484.1	489.9	86.3	484.1	737.2	40	8	600
97	184.3	327.9	61.1	340.1	334.7	61.1	355.6	653.0	40	8	400
98	194.2	261.8	115.9	277.0	258.4	115.9	315.3	648.5	40	8	320
99	507.3	720.9	86.1	725.2	795.8	86.1	738.6	914.1	40	8	960
100	230.3	150.2	86.1	151.1	110.6	86.1	174.1	418.0	40	8	200
101	190.0	37.5	86.1	37.8	27.6	86.1	43.5	230.4	40	8	50
102	287.1	81.0	93.9	102.4	101.2	93.9	180.6	371.7	40	8	100
103	202.3	34.2	76.1	34.8	36.8	76.1	138.6	247.1	40	8	50
104	134.0	13.7	76.1	13.9	14.7	76.1	84.9	169.8	40	8	20
105	132.5	248.0	54.1	265.6	301.9	54.1	251.5	530.5	40	8	320

106	258.9	77.5	54.1	83.0	48.9	54.1	137.9	326.5	40	8	100
107	182.8	38.7	54.1	41.5	25.6	54.1	87.4	239.0	40	8	50
108	218.2	126.4	74.0	138.8	71.1	74.0	172.6	432.5	40	8	200
109	95.6	137.9	31.3	120.0	176.8	31.3	120.0	210.7	20	8	160
110	94.1	68.9	31.3	86.2	109.7	31.3	100.6	174.4	20	8	80
111	207.7	298.1	25.5	225.9	302.6	25.5	225.9	306.0	20	8	480
112	75.2	61.5	30.2	53.1	77.1	30.2	81.3	145.5	20	8	100
113	55.5	94.3	34.2	94.3	94.6	34.2	97.4	164.6	15	8	120
114	46.4	58.9	34.2	58.9	59.1	34.2	61.2	127.8	15	8	75
115	101.1	33.5	29.2	34.8	30.2	29.2	47.5	113.7	20	8	60
116	80.7	84.6	48.0	92.8	90.8	48.0	82.1	204.6	20	8	100
117	13.5	83.3	30.7	73.1	85.9	30.7	81.1	182.5	20	8	150
118	29.6	101.3	32.8	111.8	114.1	32.8	131.0	234.6	20	8	200
119	70.0	193.9	22.8	141.1	191.0	22.8	141.1	217.4	20	8	300
120	16.9	97.5	25.6	111.1	108.5	25.6	133.8	207.3	20	8	160
121	88.3	75.6	23.0	101.1	88.0	23.0	143.2	185.4	20	8	160
122	124.7	149.0	14.4	181.7	125.6	14.4	188.0	211.8	20	8	300
123	128.2	149.8	47.1	192.3	172.5	47.1	228.0	290.8	20	8	300
124	60.3	83.7	16.3	97.5	90.2	16.3	115.0	161.5	20	8	160
125	156.6	175.4	20.3	199.0	172.6	20.3	232.2	260.5	20	8	300
126	5.0	17.5	4.8	12.7	17.5	4.8	12.7	17.5	5	8	50