

Niestety, w pewnych sytuacjach system SVN okazuje się niewystarczający. Przykładowo, pojedynczemu programiście zdarza się pracować równolegle nad kilkoma zmianami, np. nad dwoma drobnymi usprawnieniami w module obrazków w edytorze *Text*. Innym problemem jest konieczność stałej łączności z serwerem przy wykonywaniu praktycznie każdej operacji związanej z repozytorium. Istnieją inne systemy kontroli wersji, w których podane problemy nie występują. Jednym z nich jest system Git, w którym każdy programista może tworzyć odgałęzienia na swoim komputerze lokalnym bez konieczności nawiązania połączenia sieciowego z centralnym repozytorium. Co więcej, po zakończeniu prac zmiany te mogą być dokładane pojedynczo do głównej linii rozwoju projektu bez tworzenia w niej rozgałęzień (zapobiega to nadmiernemu bałaganowi w grafie zmian repozytorium).

W systemie Git ciekawe są nie tylko funkcje wspierające w pełni rozproszony system pracy, lecz także sama historia jego powstania. Początkowym przeznaczeniem Gita było przechowywanie kodu jądra systemu operacyjnego Linux, a powstał on wtedy, gdy wcześniej używany do tego system kontroli wersji o nazwie BitKeeper przestał być darmowy dla projektów o otwartym kodzie źródłowym. Prace nad Gitem rozpoczęły się 3 kwietnia 2005 roku, a już dwa i pół miesiąca później, 16 czerwca, system był używany do przechowywania kodu jądra Linuksa. Warto odnotować, że zgodnie z wolą twórców, do dziś Git stanowi tzw. wolne oprogramowanie, co generalnie oznacza, że każdy może go za darmo używać i modyfikować praktycznie bez ograniczeń. System Git, jakkolwiek bardzo funkcjonalny, jest koncepcyjnie nieco trudniejszy do ogarnięcia od „mniej rozproszonych” systemów takich jak SVN.

W tym artykule staraliśmy się pokazać, dlaczego każda większa firma z branży komputerowej używa systemu kontroli wersji. Na koniec warto jednak zapytać, czy takie systemy mogą się przydać „zwykłemu śmiertelnikowi”, który nie jest akurat dyrektorem ani pracownikiem takiej korporacji. Otóż tak! Systemy kontroli wersji świetnie spisują się w sytuacji, gdy kilka osób wspólnie pracuje nad jednym projektem, na przykład pracą naukową, redakcją książki czy niewielkim programem komputerowym. Ale nie tylko! Nawet jedna osoba może z powodzeniem stosować systemy takie jak SVN do przechowywania ważnych danych na komputerze. W sieci dostępnych jest wiele darmowych i prostych w obsłudze serwisów udostępniających serwery na repozytoria SVN. Dobrą motywacją jest przypomnieć sobie, kiedy ostatnio usunęło się lub przypadkowo nadpisało jakiś bardzo ważny plik...



Reklama w Internecie, czyli o aukcjach Vickreya

*doktorant, Wydział Matematyki,
Informatyki i Mechaniki,
Uniwersytet Warszawski

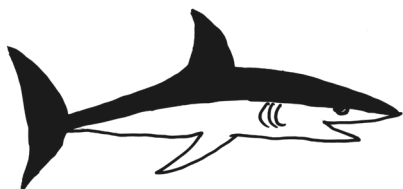
*Karol KURACH**

Czy zastanawiałeś się kiedykolwiek, drogi Czytelniku, na czym zarabiają największe firmy informatyczne w Dolinie Krzemowej? Google nie pobiera opłat za wyszukiwanie informacji, konto na Facebooku jest darmowe, Yahoo nie żąda pieniędzy za e-mail... Z punktu widzenia użytkownika praktycznie wszystko wydaje się bezpłatne. Dlaczego zatem powyższe firmy wyceniane są na miliardy dolarów? Kto pokrywa koszty utrzymania serwerów, zużycia prądu czy pensji pracowników?

Kluczem do odpowiedzi na postawione pytania są reklamy internetowe. Zyski z nich stanowią główne źródło dochodów każdej z powyższych firm (np. w przypadku Google w roku 2010 było to 96% całkowitego dochodu, a w przypadku Yahoo około 90%). Jest to bardzo dynamicznie rozwijająca się branża – wydatki przeznaczane na ten typ reklamy rosną co roku o miliardy dolarów. W samych tylko Stanach Zjednoczonych było to ponad 30 mld dolarów w roku 2010, z szacowanymi 50 mld dolarów na rok 2015. W artykule postaram się przybliżyć najważniejsze pojęcia i mechanizmy związane z tym zagadnieniem.

Co mają wspólnego reklamy z aukcjami? Wpiszmy w wyszukiwarce internetowej jakieś zapytanie. Otrzymamy listę wyników i zapewne również pewien zbiór odnośników wyróżnionych jako *sponsorowane*. Te odnośniki to przykład tzw. *marketingu bezpośredniego* (ang. *direct marketing*). Są one nastawione na natychmiastową akcję ze strony użytkownika – kliknięcie i kupienie czegoś. To, jakie reklamy zostaną wyświetlone, zależy od treści zapytania. Takie podejście ma za zadanie zmaksymalizować skuteczność reklamy oraz zmniejszyć ilość spamu docierającego do odbiorców.

Wyświetlane reklamy zależą również od innych czynników, jak np. adres IP, kraj, z którego pochodzi zapytanie, pora dnia itp. Dla uproszczenia pominiemy te czynniki w naszych rozważaniach.



Miejsca na reklamy są sprzedawane na aukcjach. Za każdym razem, gdy użytkownik wyświetla stronę, przeprowadzana jest aukcja, która decyduje, jakie reklamy zostaną pokazane (temat ten został opisany szerzej w dalszej części artykułu). Opłata pobierana od reklamodawcy zależy od modelu płatności, który został wybrany. Trzy główne modele płatności występujące obecnie to:

- *CPM* (ang. *cost per mile*) – reklamodawca płaci x za każde 1000 wyświetleń reklamy,
- *CPC* (ang. *cost per click*) – reklamodawca płaci y za każde kliknięcie w reklamę (czyli przejście na stronę internetową reklamodawcy),
- *CPA* (ang. *cost per action*) – reklamodawca płaci z za każdą *akcję* wykonaną przez użytkownika. *Akcją* może być np. kupienie czegoś w sklepie internetowym.

Intuicyjnie, dla ustalonej reklamy zachodzi nierówność $\frac{x}{1000} < y < z$. Aktualnym standardem wykorzystywanym przez takie firmy jak Google czy Facebook jest *CPC*. *CPM* używany jest głównie do płacenia za wysyłanie reklam pocztą elektroniczną oraz za reklamy wyświetlające obrazy lub bannery (tzw. *display ads*). *CPA*, pomimo ogromnej wygody dla reklamodawców (płacą tylko za efekty), ma pewne istotne wady. Często trudno rozstrzygnąć, czy dana akcja rzeczywiście została wykonana, czy też nie – sprzedawca mógłby zaniżyć zgłaszaną ilość akcji, aby zapłacić mniej. Dodatkowo, o wiele mniej osób wykonuje akcje niż klika w linki. Ponieważ *CPA* generuje o wiele mniej danych niż *CPC*, trudniej w tym modelu dobrze zaplanować strategię wyświetlania reklam. W dalszych rozważaniach skupimy się dla uproszczenia tylko na modelu *CPC*.

Które reklamy się wyświetlają? Dla każdej reklamy klient określa związane z nią słowa kluczowe (oznaczane dalej przez s_i) oraz maksymalny *CPC*, który jest skłonny zapłacić. Przykładowo, jeśli chcielibyśmy zareklamować własną firmę zajmującą się udzielaniem korepetycji z matematyki, takimi słowami (frazami) mogłyby być $s_1 =$ „korepetycje”, $s_2 =$ „matematyka”, $s_3 =$ „przygotowanie do matury”. Za każdym razem, gdy ktoś wpisałby zapytanie tematycznie powiązane z co najmniej jedną frazą s_i , nasza reklama byłaby *aktywna*, tzn. spełniałaby warunek konieczny do wyświetlenia jej.

Ponieważ jednak wiele reklam może zawierać dane słowo kluczowe (nasi konkurenci z branży korepetycji również się reklamują), a liczba miejsc na nie jest ograniczona, powstaje problem – które reklamy wybrać? Istotna jest również ich kolejność – im wyżej w rankingu znajduje się reklama, tym częściej będzie ona klikana. Ta właściwość opisywana jest parametrem *CTR* (ang. *clickthrough rate*), który oznacza stosunek liczby kliknięć do liczby wyświetleń. Niech CTR_i oznacza wartość *CTR* dla i -tego od góry miejsca na reklamę. Zakładamy, że zachodzi $CTR_i \geq CTR_{i+1}$. Wartość CTR_i można w przybliżeniu rozumieć jako „prawdopodobieństwo, że reklama ustawiona na pozycji i zostanie kliknięta”.

Oczekiwany zysk wyszukiwarki z wyświetlenia reklam wynosi zatem

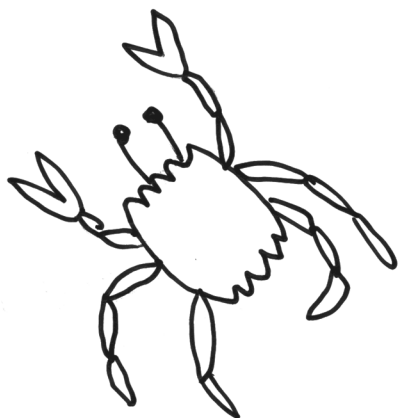
$$\sum_i CTR_i \cdot CPC_i$$

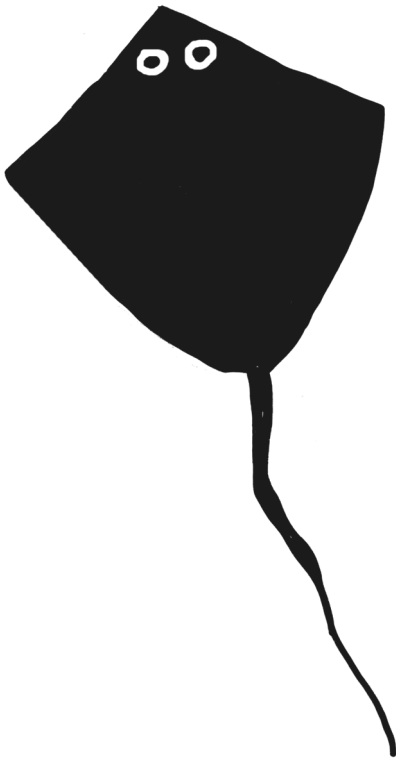
gdzie CPC_i oznacza *CPC* dla reklamy przydzielonej do miejsca i . Zauważmy, że funkcja ta osiąga maksimum, gdy posortujemy reklamy malejąco ze względu na *CPC*, i w takiej kolejności przydzielimy im miejsca. Prosty dowód tego faktu pozostawiamy Czytelnikowi.

Koszt reklamy i problem braku stabilności. Użytkownik wyszukiwarki wpisał zapytanie, nasza reklama wyświetliła się i została kliknięta – ile za to zapłacimy? Najprostszym rozwiązaniem, które przychodzi do głowy, jest „niech reklamodawca płaci zawsze tyle, ile zadeklarował”. Czyli jeśli ktoś ustawił $CPC = 1,0$, to płaci 1 zł, a jeśli $CPC = 0,5$, to 50 groszy. To podejście wydaje się dosyć intuicyjne, i rzeczywiście przez wiele lat było stosowane do ustalania opłat.

Takie rozwiązanie ma jednak pewną istotną wadę, którą przedstawimy na przykładzie. Załóżmy, że dane są dwie reklamy r_1 i r_2 oraz $CPC_{r_1} = 0,02$, $CPC_{r_2} = 0,01$. W takiej aukcji wygrywa r_1 i dostaje pierwsze od góry miejsce na reklamę. W tym momencie właściciel r_2 może pomyśleć: „pierwsze miejsce jest o wiele bardziej wartościowe – zapłacę dodatkowe 2 grosze, aby je zdobyć”. Po zmianie CPC_{r_2} na 0,03 druga reklama zacznie wygrywać. Tak samo pomyśli więc właściciel r_1 i ustawi $CPC_{r_1} = 0,04$.

W naszym modelu *CTR* jest zależny jedynie od pozycji reklamy na liście wyników. W rzeczywistości branż jest pod uwagę więcej czynników, w tym również zmieniających się w czasie (np. jakość reklam – niektóre mogą być bardziej interesujące i chętniej wybierane przez użytkowników). Dlatego też nie zawsze klient z najwyższym *CPC* dostanie najlepsze miejsce reklamowe.





Ta „gra” między reklamodawcami będzie trwała aż do pewnej kwoty X , przy której posiadacz drugiego miejsca (powiedzmy, że r_2) nie będzie chciał przebić oferty. Może on wtedy pomyśleć: „skoro i tak mam drugie miejsce, to przynajmniej zmniejszę moje koszty” i ustawić z powrotem $CPC_{r_2} = 0,01$. Co, oczywiście, skutkuje tym, że r_1 wkrótce zmieni swoje CPC na $0,02$, bo taka cena wystarcza, aby utrzymać pierwszą pozycję. Wróciliśmy zatem do punktu wyjścia i gra zacznie się wkrótce na nowo...

Brak „optymalnej strategii”, konieczność ciągłego pilnowania ceny oraz uczucie „przeplacania/bycia oszukiwanym” były poważnymi mankamentami tej metody i stanowiły motywację do poszukiwania lepszych rozwiązań.

Aukcja Vickreya. Rozpatrzmy aukcję, na której dostępny jest tylko jeden niepodzielny przedmiot i n osób chciałoby go kupić. Niech w_i oznacza rzeczywistą wartość przedmiotu dla i -tej osoby, a l_i – kwotę, którą zaliczyła i -ta osoba. Przedmiot dostanie osoba, która zaoferowała najwięcej, i będzie musiała zapłacić za niego tyle, ile wyniosła druga najwyższa oferta. Taka aukcja nazywana jest również „aukcją drugiej ceny”. Została ona wymyślona w 1961 roku przez Williama Vickreya, profesora Uniwersytetu Columbia w Nowym Jorku i laureata Nagrody Nobla z ekonomii. Okazuje się, że w tak zdefiniowanej aukcji dla każdego gracza najlepszym rozwiązaniem jest zaliczyć dokładnie tyle, ile przedmiot jest dla niego wart (w_i).

Twierdzenie 1. Rozpatrzmy aukcję Vickreya dla n , w_i , l_i opisanych wyżej. Dla i -tego gracza zaliczenie rzeczywistej wartości przedmiotu ($l_i = w_i$) jest strategią optymalną.

Dowód. Zysk i -tego gracza w aukcji wynosi:

$$\text{zysk}(i) = \begin{cases} w_i - \max_{j \neq i} l_j & \text{jeśli } l_i > \max_{j \neq i} l_j, \\ 0 & \text{w przeciwnym przypadku.} \end{cases}$$

Niech $m_i = \max_{j \neq i} l_j$. Przez S będziemy oznaczali strategię polegającą na ustaleniu $l_i = w_i$.

Rozpatrzmy strategię, w której $l_i > w_i$, czyli gdy gracz licytuje więcej, niż przedmiot jest dla niego wart. Mamy trzy podprzypadki:

- Jeśli $m_i > l_i$, to gracz nie wygrał aukcji, więc zysk wynosi 0. Tyle samo wyniosłby zysk przy wyborze strategii S , wtedy również gracz nie wygrywa aukcji.
- Jeśli $m_i \leq w_i$, to gracz wygrał aukcję i zarobił $w_i - m_i$. Przy wyborze strategii S zarobiłby jednak dokładnie tyle samo.
- Jeśli $w_i < m_i \leq l_i$, to gracz wygrał aukcję, ale w porównaniu do S stracił na zakupie $m_i - w_i$.

We wszystkich podprzypadkach strategia S okazała się nie gorsza od licytowania więcej niż w_i . Analogicznie rozpatrujemy strategię, w której $l_i < w_i$, czyli gdy gracz licytuje mniej, niż przedmiot jest dla niego wart. Jeśli $m_i > w_i$, to zarówno w tej strategii, jak i w S , gracz nie wygrywa aukcji. Jeśli $m_i \leq l_i$, to gracz w obu przypadkach wygrywa i zarabia $w_i - m_i$. A jeśli $l_i < m_i \leq w_i$, to zyski ze strategii są różne – gracz zaliczył mniej niż m_i , więc traci przedmiot warty w_i , osiągając zysk 0. W strategii S gracz kupiłby przedmiot z zyskiem $w_i - m_i$.

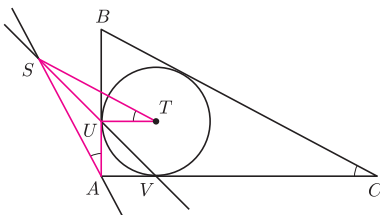
Za każdym razem zysk gracza nie był większy niż w S , więc S jest strategią optymalną. \square

W przypadku reklam internetowych licytowanym przedmiotem jest miejsce na reklamę. Liczbie l_i odpowiada CPC_i , natomiast w_i to wartość, jaką ma dla nas kliknięcie w reklamę. Jak wyznaczyć w_i ? Można próbować oszacować tę wartość na podstawie danych historycznych. Jeśli obliczymy zyski ze sprzedaży towarów użytkownikom, którzy weszli przez link reklamowy, i podzielimy je przez liczbę takich wejść, dostaniemy „oczekiwany zysk” z jednego kliknięcia.

Aukcja Vickreya rozwiązuje postawiony problem, ale pamiętajmy o ograniczeniu – tylko jeden przedmiot, czyli jedno miejsce na reklamę. W rzeczywistości wyszukiwarki internetowe oraz portale społecznościowe wyświetlają zazwyczaj po kilka reklam. Polecamy Czytelnikowi zastanowienie się nad problemem uogólnienia aukcji Vickreya na przypadek więcej niż jednego przedmiotu.



Rozwiązanie zadania M 1354. Zauważmy, że trójkąty AUS, TUS są przystające na mocy cechy bkb (oczywiście $TU = AU$, a ponadto $\sphericalangle AUS = 180^\circ - \sphericalangle AUV = 135^\circ$, $\sphericalangle TUS = 180^\circ - \sphericalangle TUV = 135^\circ$).



Zatem $\sphericalangle UTS = \sphericalangle UAS$. Natomiast z twierdzenia o stycznej do okręgu i kącie wpisanym, zastosowanego do okręgu opisanego na trójkącie ABC , mamy $\sphericalangle UAS = \sphericalangle ACB$. Wobec tego kąty UTS i ACB są równe, co kończy dowód.