Competition in Search Engine Market

Wugang Zhao Stanford University • Stanford, CA Edison Tse Stanford University • Stanford, CA

Abstract

Google has been very successful in becoming the dominant search engine platform in the US and Western countries. In contrast to its success in the US and Western countries, Google does not have much luck in some Asian countries such as China and Korea. Google is far behind in terms of search market share in China and Korea. In this paper, we model the search engine market as a two-sided markets model and analyze the industry structure and competition of the search engine market. First, we present a mathematical model for a general search engine two-sided market. Then we use the model to analyze the search engine history and explain why multiple search engines could co-exist in the early days of the search engine history. We also explain how Google, a latecomer in the search market, could become the leading search engine, and how Google has strengthened its leading position. Next we apply the model to China and Korea's search markets and analyze how Google could lose the game to local search companies. In the end, we propose some strategies on how search engine market leaders could maintain and strengthen their leading positions.

Introduction

The World Wide Web has experienced exponential growth since its inception in early 1990s, and it continues growing rapidly (W3C, 2000). Today, it is used in almost every aspect of our everyday life, such as searching for online shopping sites, company and product information sites, online social networking sites and online news sites. Google was estimated to index around 20 billion web pages in Oct. 2010 (Kunder, 2010). Netscraft reported that the total number of websites reached 232 million in Oct. 2010 (Netcraft, 2010). These data indicate that there are huge amount of information out there on the Internet, with millions of websites and billions of web pages.

Search engines hold the key to the wealth of information available on the Internet. Users spend a significant amount of time on search engines looking for relevant information (Gandal, 2001). Very often a user starts with a search engine to get to a desired website. For example, an online consumer who wants to buy a digital camera may go to Google, search for digital camera, browse the products and reviews of the candidate sites, and pick one to make the purchase, or a researcher

who is looking for a paper may go to a search engine, search for the title of the paper, and download the paper from a target site. Search engines play a critical role in disseminating web site information on the Internet and they have a powerful position in the web space. According to Alexa, search engines are among the most actively visited sites on the Internet. For example, Google ranks #1 in the most visited site in 2010 and several other search engines are among the top 25 (Alexa, 2010). With the central role search engines are playing, they have become very attractive options for online advertising and target marketing.

In the early days of the search engine history, there were many different search engines in the market. Google's entrance into the search engine market changed the market structure and competition landscape. Google, a latecomer in the search engine market, was founded in 1998 and rose to prominence for its advanced search technology in 2000. Google surpassed Yahoo as the leading search engine in 2002 and has strengthened its market leadership position since then (SearchEngineWatch; NETMARKETSHARE, 2010). Google also expanded into international search markets after it established its lead position in the US. Compared to its huge success in the US and Western countries, Google did not have much luck in some of the major Asian countries such as China and Korea. In China, which has the largest Internet population and potentially the largest Internet market in the world, Google is far behind the local search market leader Baidu. Baidu holds around 70% of the search market while Google has only around 21%. In Korea, the local market leader Naver has 62% search market share and Google only has 4% (Bonfils, 2010; MarketThe-Globe, 2010; VisualEconomics, 2010).

In this paper, we will model the search engine market as a two-sided markets model and derive some insights from it. The two-sided markets model has attracted a lot of interest among researchers (Armstrong, 2004, 2005; Csorba & Hahn, 2003; Eisenmann, et al. 2006; Ferrando et al. 2004; Rochet & Tirole, 2003, 2004, 2006; Roson, 2005; Rysman, 2000). We will use a similar two-sided markets model presented by Zhao and Tse (2011) to model the search engine market.

Based on the model, we come up with some general strategies for a search engine to compete in the market. We will then use different variations of this model to explain the history and industry structure of the search engine market. We will explain why searchers tended to use multiple search engines to conduct the same search query in the early days of the search engine history and how Google could emerge as a leading search engine.

We then use the model to analyze Google's failure in China and Korea. We point out that Chinese users are in a different market segment from users in the US and Western countries and Chinese language contents and related services are more relevant to them. Focusing on Chinese users and offering more Chinese language contents and related services gave Baidu an edge when competing with Google in China. In Korea, the lack of Korean language contents created a special market condition. The local company Naver took advantage on this by creating a knowledge sharing service and keeping it closed. This enabled Naver to accumulate large amount of content in its private database, which Google does not have access to. This gives Naver an advantage over Google in the Korean search market. Finally, we will propose some strategies for Google and local search engine market leaders on how they could keep and strengthen their lead positions.

Search engines have attracted many research activities. Lawrence, Giles, Bradlow and Schmittlein pointed out that each search engine only indexed a small portion of the whole web space, and different search engines indexed a different set of web pages (Bradlow & Schmittlein, 1999; Lawrence & Giles, 1998, 1999). Combining the results from multiple search engines can greatly improve the coverage, as was done with meta search engine such as MetaCrawler (www.metacrawler.com). A searcher would switch to another search engine if he/she fails to find the information he/she is looking for from one search engine. A user will do this because the low switching cost. Switching cost for search engine is low due to two major reasons. First, search engine service is free for search users and search engine companies make money from advertisers. The revenue they can make from advertisers depends on the number of users who visit the particular web site (Chatterjee & Novak, 1995; Hoffman & Novak, 1996). Second, search engines usually have the same user interfaces and the knowledge in using one search engine can be easily applied to another search engine. In another paper published by Lawrence and Giles in 1999, they expanded the research and found out that search engines did not index sites equally, they indexed a biased sample of the web, and they may not index new pages for months (Lawrence & Giles, 1999).

The brand effect and first mover advantage have also been investigated in search engine competition literature. Mukhopadhyay et al. pointed out that the incumbent search engine has a first-mover advantage, and, unless the new entrant has a cost advantage, the incumbent will emerge as the leader in equilibrium (Mukhopadhyay, Rajan, & Telang, 2002; Telang, Rajan, & Mukhopadhyay, 2004). Telang et al. (20004) found that users develop loyalty for a given search engine. But simple search users do not develop strong loyalty. They showed that good quality search results are essential for repeated engine use. A poor quality engine cannot hope to develop a loyal base (Telang, Mukhopadhyay, & Wilcox, 2000; Telang & Mukho-

padhyay, 2005). Gandal (2001) examined the evolution of and competition in the Internet search engine market and found that while early entrants (Yahoo, Lycos, Excite, Infoseek, and Altavista) still have an advantage, the pure brand effect advantage has been declining over time (Gandal, 2001). Jansen, Zhang, and Zhang (2007) investigated the effect of brand awareness on search engine competition results. They found that even though different search engines have similar technology and similar user interfaces, the majority of search traffic is directed to a small number of search engines. In this paper, we propose some strategies that help a first mover keep its leadership position. One important strategy is to build a new two-sided markets platform with strong positive cross network effect by leveraging its existing customer base.

The paper is organized as follows. We first introduce the mathematical model and the key players in a search engine two-sided market in *Mathematical Model*. Then we use the model to analyze the search engine market in *Analysis of the Search Engine's History of Development*. We analyze why searchers tended to use multiple search engines to conduct a search query in the early stage of the search engine history. Then we explain how Google, a latecomer in the search engine market, could emerge as a market leader. In *Analysis of Google's Failure in China and Korea*, we analyze why Google failed in China and Korea. In *Proposed Strategies for Market Leaders*, we come up with some strategies for local search leaders and Google based on our model analysis and Google's failure in China and Korea. *Conclusion and Discussion* summarizes and discusses the limitation of our research as well as the potential future research.

Mathematical Model

A search engine connects searchers (information seekers) and web sites (information providers). The web sites contain different kinds of information. For example, a company's web site that has the product's information, or an online catalog that has many research papers and books. Searchers try to find the information from the web sites. Due to the huge amount of information available on the Internet, it's hard for a searcher to quickly find the information he/she is looking for without the help of a search engine. A search engine stores and indexes the web pages so it can help a searcher quickly find the information.

A search engine creates a two-sided markets platform, on one side are the searchers (information seekers), and on the other side are the web sites (information providers). There are positive cross network effects between the two sides: The more

searchers that use a search engine to search for information, the more benefit for the web sites owners (since the information published to the web sites can be viewed by more people); the more web sites publish information on the Internet and these pages are indexed by a search engine, the more benefit to the searchers since they are more likely to find the information they are looking for. Figure 1 shows a diagram of a general search engine two-sided markets platform. In this diagram, *S* stands for the whole web space, the universe. *"Web Sites"* on the right hand side are the web pages indexed by the search engine. *"Searchers"* on the left hand side are the users who come to the search engine platform to conduct the search queries. *"+"* signs refer to the positive cross network effect between the searchers and web sites.

Figure 1 A General Search Engine Two-sided Markets Platform



The two-sided markets platform a search engine creates possesses some unique characteristics. First, a search engine has access to all the web sites that are publicly accessible on the Internet. So a search engine platform does not face the chicken-and-egg problem faced by a general two-sided markets platform since a search engine can easily index a large set of web pages to build one side of the platform. Second, search engines offer the service free of charge for both the searchers and web sites because they make money from advertisers. The users that are attracted to a search engine portal are the most valuable assets to the search engine platform. This revenue model is similar to that of the television network (Beebe, 1977; Steiner, 1952). Figure 2 shows the diagram of a general search engine two-sided markets platform with advertisers. In this diagram, we can see that the searchers have positive cross network effect on the advertisers since the more searchers come to a search engine platform, the more likely the advertisers will get sales through advertising on the search engine platform. On the other hand, the advertisers have negative cross network effect on the searchers because the advertisements on the search engine may distort the search results or disturb the searchers. Usually search engines display advertisements separately on the right hand side of the search results to minimize the negative impact on the searchers.

Figure 2 A General Search Engine Two-sided Markets Platform with Advertisers



We can use a similar two-sided markets model as described by Zhao and Tse (2011). Let U be the utility a searcher derives from using the search engine platform, NB be the net benefit, and C be the cost of using the service.

The utility a searcher derives from using the search engine platform depends on the quality of the search platform itself, the amount of web pages indexed, and how relevant the search result pages are. The quality of the search platform includes things like the usability of the site, response time of processing a user's query, etc. We will use a single variable Q to denote the comprehensive quality of the search platform. Comprehensive quality captures all the factors that determine the quality of the platform. The amount of web pages determines how likely the user will get the results he/she is looking for. The more web pages indexed by the search platform, the more likely the user will find the information. We will use M to denote the set of pages the search platform indexes. As shown in Figures 1 and 2, S refers to the total set of web pages on the Internet. Another critical factor is the capability of the search platform to find the relevant web pages the user is looking for from the huge amount of web pages it indexes. We will call this relevance factor, and denote it as *R*. The greater *R* is, the better the search engine. *R* is determined by the search technology and it is a critical factor in determining how much value a searcher derives from using a search engine. Even if a search engine platform indexes huge amount of web pages (large |M|), if the relevance factor R is low, it won't offer much value to the searchers, since the page a searcher is looking for will be buried in a lot of irrelevant pages.

With the above definitions and notations, we have:

$$U_{searcher} = (1 - \gamma)f(Q) + \gamma Rg(|M|). \text{ Here } \gamma \text{ is a constant, and } \gamma \in [0,1]$$
(2.1)

R is positive. |M| is the cardinality of set M and the size of the indexed web pages. A searcher's value comes from two sources: the search engine platform itself and the indexed web pages. γ is a parameter to denote the percentage of a searcher's value that comes from the indexed web pages. Since a searcher comes to a search engine platform to look for web sites or pages with the information he/she is looking for, the majority of a searcher's value comes from the indexed web pages and how relevant the returned results are. The quality of the search engine platform is important, but not as important as the indexed web pages and how relevant the returned results are. No matter how good the U_i of the platform is and how fast the search engine can return the results, if it does not have the page the searcher is looking for or cannot return the relevant page, there is not much value to the searcher. So is large (between 0.5 and 1. In general, indicates the strength of the positive cross network effect between the two sides of a two-sided markets model. The larger is, the stronger the positive cross network effect). f and g are monotonic increasing functions. To simplify the analysis, we can assume that f(Q) is in proportion to Q, and g(|M|) is in proportion to |M|, we can write U as:

$$U_{searcher} = (I - \gamma)\alpha Q + \gamma \beta R |M|. \ \gamma \text{ is a constant, and } \gamma \in [0,1], \ \alpha, \beta \text{ are}$$

positive constants (2.2)

Let C be the cost incurred from using a search platform. The cost of using a search engine platform is determined by the time spent searching and browsing the search results (since the search platform is free to searchers in real world). We have

 $NB_{searcher} = (1 - \gamma)\alpha Q + \gamma \beta R|M| - C. \gamma$ is a constant, and $\gamma \in [0,1], \alpha, \beta$ are positive constants

Rappa (2009) listed four key elements that a searcher uses to pick a search platform: Relevance, Comprehensiveness, Freshness and Speed. We capture the relevance through relevance factor R, comprehensiveness through M. Freshness and speed are captured by comprehensive quality Q.

The searchers have positive cross network effect on the advertisers. Advertisers come to a search platform since their advertisements can be viewed and clicked by searchers and potentially be converted into purchases. The value an advertiser derives from joining a search platform depends on three key factors: the number of searchers, the number of advertisers, and the relevance factor. The more searchers on the search platform, the more likely the advertisement will be viewed and clicked; the more advertisers, the less likely the searchers will see the advertisement from an individual advertiser; the larger the relevance factor, the more likely the search platform will display the advertisement that interests the searchers. We use X to denote the number of advertisers. So we have

$$U_{advertiser} = \delta R \frac{X}{K}, \ \delta \ is \ a \ positive \ constant$$
 (2.4)

Here we assume $U_{advertiser}$ is a linear function of $R\frac{X}{K}$, δ is the coefficient. Let *P* be the price an advertiser pays to the search platform, we have

$$NB_{advertiser} = \delta R \frac{X}{K} - P, \ \delta \ is \ a \ positive \ constant$$
(2.5)

If we consider all the advertisers as a whole, we have

$$U_{advertisers} = \delta RX, \ \delta \ is \ a \ positive \ constant \tag{2.6}$$

Let P_{total} be the total price advertisers pay to the search platform, we have

$$NB_{advertisers} = \delta RX - P_{total}, \delta is a positive constant$$
 (2.7)

 $U_{advertisers}$ and $NB_{advertisers}$ are the total utilities and net benefits for all the advertisers on the search engine platform, respectively.

Based on the above model, a search engine platform can adopt the following strategies to attract more searchers to use its service.

- Increase the number of web pages indexed, this will increase M.
- Improve on the relevance of the search results, to improve the relevance factor R
- Increase the quality of the search engine platform itself, such as improve on usability, reduce response time for processing a user's query, etc.

- Consolidate search results from several search engines to increase the coverage, e.g., MetaSearch.
- Merge with and acquire other search companies to increase the coverage of the total indexed web pages.
- Reduce the negative impact of advertisements on the search results and searchers.
- Offer free service to searchers and make money from other sources.

Analysis of the Search Engine's History of Development The Early Stage

Today, there are only a few major search engines in the market. From U.S. search engine market share data in June 2011, Google was in a dominant position with 84.58% market share, Yahoo had 8.13%, Bing had 5.38%, and the rest had no more than a 1% market share (KARMA SNACK, 2011). In the early stages of the search engine history, the market was very fragmented. There were many search engine platforms in the market, many of which no longer exist today. Due to the limitation of technologies, each search engine only covered a small portion of the total web space, and searchers may have used multiple search engine platforms during one query session (Bradlow & Schmittlein, 1999; Lawrence & Giles, 1999). For example, in 1997, Yahoo led the market with around 34% market share, Infoseek had 18%, Excite had 17.6%, Lycos had 11.4%, Altavista had 10.9%, Webcrawler had 7.4%, and there were many of others with smaller market shares (Gandal, 2001).

To explain this with the model we built in Figure 2, let's assume that there are two search engine platforms: SE1 and SE2 and they have same quality Q and same relevance factor R. Let the sets of web pages indexed by these two search engines be . If the searchers only use one of the search engines, as shown in Figure 3, the net benefits are as follows:

 $NB_{I} = (1 - \gamma)\alpha Q + \gamma \beta R|M_{I}| - C_{I}$ $NB_{2} = (1 - \gamma)\alpha Q + \gamma \beta R|M_{2}| - C_{2}$



Figure 3 Searchers Only Use One Search Platform

If the searchers use both search engines for their queries at the same time, as shown in Figure 4, the net benefit will be

$$NB = (1 - \gamma)\alpha Q + \gamma \beta R|M| - C_1 - C_2.$$

So $NB - NB_1 = \gamma \beta R(|M| - |M_1|) - C_2 = \gamma \beta R(|M_1 \cup M_2| - |M_1|) - C_2.$

When M_1 and M_2 each is a small subset of S, and there is not much overlap between them, $|M_1 \cup M_2| \approx |M_1| + |M_2|$. So $NB - NB_1 \approx \gamma \beta R |M_2| - C_2$. Since most of a searcher's value comes from the indexed web pages, $\gamma \beta R |M_2| - C_2 > 0$. So we have $NB - NB_1 > 0$, similarly we have $NB - NB_2 > 0$. In this case, searchers would use both search engines to conduct the same search query.

When M_1 and M_2 is a large subset of S, we have $|M_1| \approx |M_2| \approx |M_1 \cup M_2| \approx |S|$. So $NB - NB_1 \approx C_2 < 0$. Similarly we have $NB - NB_2 < 0$. In this case, searchers would only use one search engine to conduct the search query.

The same deduction applies to the case with more than two search engines, and the case in which different engines have different quality and relevance factors. So in the early stages of search engine history when the coverage of each individual search engine was low compared to the whole web space, the searchers could derive more value by using multiple search engines at the same time. Also, a MetaSearch engine could offer greater values to searchers by combining the search results from several other search engines. The entry barrier was also low due to the limited coverage of search engines in the market. A new search engine could bring additional value to searchers since it could improve on the coverage of the indexed web pages.

Based on the above analysis, we can tell that the searchers were multi-homing since they tended to use multiple search engines at the same time in the early stage of the search engine history. With the advancement of technology, search engines could greatly improve on the search relevance and the sets of indexed websites. M_1 and M_2 can then be increased to close to S, so $|M_1| \approx |M_2| \approx |M_1 \cup M_2| \approx |S|$. In this case, $NB - NB_1 < 0$ and $NB - NB_2 < 0$. In this case, searchers would use one search engine to conduct a search query. So the tendency of multi-homing is low when the each search engine can cover most of the web.





How Google Became Successful and Why it's so Powerful

Google is a latecomer in the search engine market. It was founded in 1998 and rose to prominence in 2000 for its advanced technology. Google invented PageRank, an algorithm used by its search engine to assign a numerical weighting to a web page to indicate the importance of that page in the web space (Hammonds, 2003). PageRank greatly improved the relevance factor R of Google's search engine and this greatly improved the relevance of the search results. The improvement in search relevance also helped Google to display the advertisements to the searchers that they are likely interested in. Google also invested massively in the search infrastructure. It built its own servers from components it bought directly from their manufacturers. According to Drummond, Google now operates the world's largest distributed computer system (VentureBlog, 2010). These investments enabled Google to index and store a much larger set of web pages than its competitors could. So Google greatly

increased the size of M, and thus increased |M|/|S|. The technology investments also enabled Google to process search queries much more quickly. It was reported that an average search query used to take 3 seconds, and Google optimized it to 0.2 second (Hammonds, 2003). Compared to the search engine platforms on the market, Google's search engine platform had better comprehensive quality: it had a very nice and neat user interface with a single search box, it could process a search query very quickly and show the results in no time, it kept the indexed pages up to date, and it showed the advertisements to the right hand side of the search results so the searchers won't be disturbed by advertisement. Google's search engine also had a much larger relevance factor R, and larger set of indexed web pages M. All these factors enabled Google to offer a much larger net benefit to searchers, and searchers quickly moved to Google and chose Google as their favorite search engine. In 2002, Google surpassed Yahoo and became the No.1 search engine in the US (Hitwise).

Google has experienced exponential growth in the past ten years and it has been successful in maintaining and strengthening its leadership position. In 2010, Google had 72% search market share in the US. Google has also expanded into other countries and has been successful. Except for a few countries, Google is the leading search engine platform across the world (VisualEconomics, 2010).

Can Other Search Engines Surpass Google?

Zhao and Tse (2011) pointed out several key advantages of the two-sided markets model compared to the value chain model. One of the key advantages of the two-sided markets model over the value chain model is that the positive cross network effect between the two sides of the two-sided markets model can help the market leader to keep its lead position. It's a lot harder, if not impossible, for a new entrant to a two-sided market to catch up and surpass the market leader (Zhao & Tse, 2011).

As we have described earlier, the general search engine business model is a two-sided markets model, but a special one. The specialness lies in the fact that every search engine has access to all the publicly accessible web space. As long as a web page is publicly accessible, a search engine can index that page and make it part of the search engine's set of indexed web pages. Because of this specialness, there is no chicken-and-egg problem for a search engine platform with a good search technology. A new entrant search platform can index a large set of web pages to build one side, and attract searchers to join the platform to build the other side. So the lead search engine does not have the advantage over new entrant in terms of having a large set of indexed web pages, because the new entrant can index a comparable large set of web pages. Whether a searcher joins a search platform or not depends on the search platform's comprehensive quality and relevance factor. In this sense, a general search engine business model is similar to a value chain model. For example, if a new search engine platform can invent a new and much better search technology than Google's, it will greatly increase the comprehensive quality and relevance factor. Since it can index the same amount of web pages, it can offer searchers much greater net benefits than Google can. Since the switching cost is zero, the searchers will move away from Google to join this new search platform and Google will lose its market lead position. Once the searchers have moved away from Google, advertisers will move away too due to the strong positive cross network effect the searchers have on the advertisers.

Based on the theory proposed by Zhao and Tse (2011), Google needs to introduce a new positive cross network effect to its search platform to help keep its lead position. This is exactly what Google has done. As shown in Figure 5, Google has attracted partner web sites to join its platform. The partner web sites are independent third party web sites, and each has its own user base and has accumulated some site traffic. Since Google is the leading search engine platform and there are many advertisers on Google's platform, the partner sites can monetize their site traffic by working with Google. Through Google's platform, the partner sites can display Google advertisements to users who visit their sites. For advertisers, they benefit from partner sites since they will reach to users beyond Google's own user base. There are strong positive cross network effects between advertisers and partner sites. Google has created a win-win situation for them. Based on our theory, the strong positive cross network effect between partner sites and advertisers will help Google keep its lead position. By attracting partner sites to Google's platform, Google's customers include not only searchers who go to Google directly, but also those who visit Google's partner sites. In 2010, 31% of Google's advertising revenue was from its partner sites (Google, 2010). The strong positive cross network effect between partner sites and advertisers helps Google to maintain and strengthen its lead position and makes it a lot harder for its competitors and new entrants to surpass it.

The most recent and powerful challenger of Google is Bing, owned by Microsoft. Bing is an improved search engine compared to MSN search. It offers a similar user interface and has comparable response time to Google. So Bing and Google have similar comprehensive quality Q. As for search relevance, some people believe Bing's search results are better, some believe Google's search results are better, and some believe they are similar (Chou, 2011; Finin, 2009; Kopp, 2011; Parr, 2009; Pogue, 2009). Since it's hard to differentiate the results from Google and Bing, we can consider that Bing has comparable search relevance factor R as Google. Also, Google has larger size of indexed web pages than Bing, so Google has larger M (Kunder, 2010).

Based on our model, Google would offer larger net benefit than Bing could offer to searchers. Therefore, Bing will not be able to replace Google as the leading search engine. However, since it improved a lot over MSN search and the market share of MSN was low, Bing would be able to increase Microsoft's search market share, and this is what had happened since its release in June 2009 (McGee, 2010).

Figure 5 A Search Engine Two-sided Market Platform with Partner Web Sites



Analysis of Google's Failure in China and Korea

Google established its leadership position in the search engine market in the US in 2002 and has been successful in strengthening its lead position and expanding into many other countries outside the US. Powered with cutting-edge technology,

huge financial resources, intellectual property and a track record of success, Google was considered almost insurmountable when it entered Chinese search market in 2006. Google thought it would conquer Chinese search market the same way it did in other Western countries. Even though Google has achieved huge success in many markets outside the US, it failed to compete with the local Chinese search engine company Baidu. In 2010, Baidu held 63% market share in Chinese search market, compared to Google's 27% (MarketTheGlobe, 2010).

Baidu was founded in 2000 when the Chinese Internet was just beginning to burgeon. It successfully established its strong position by offering something that Google did not offer initially: links to pirated songs, TV shows and movies from Chinese web sites. Baidu claimed this was legal because the media files were not on its own computers. Google did not offer a similar service until 2009, when it finally introduced a free online music service with the permission of the music labels, but it has never managed to make up the lost ground (Barboza & Stone, 2010). Even after Google introduced the free music download service, the service was often problem-atic (Salibra, 2010).

Daltorio (2010) attributed Google's failure in Chinese market to its neglect of the most important rules of any business: Know your market. Google tried to impose its Western vision of the web onto the Chinese Internet users and ended up creating a mess for itself and easy profits for local competition. Its arrogance made it ignore free music downloads, the popular service that Baidu offered and Chinese Internet users valued. Also, Daltorio (2010) pointed out some facts that showed Google expected Chinese users to adapt to it, instead of adapting itself to Chinese users. For example, the US search box did not fit Chinese characters very well and Google did not bother learning that Chinese Internet users spend most of their online time on entertainment- as compared to Europeans and Americans, who use it more for work-related purposes.

In this section, we will use the two-sided markets model to model Google and Baidu's competition and explain Google's failure in Chinese search market. Figure 6 shows the competition between Google and Baidu in the Chinese local search market. In above sections, we simply assume that the indexed web pages are homogeneous and they are equally likely to be accessed by search users. This assumption is valid when we analyze a single homogeneous market in which users have the same needs and do similar search queries. But in reality, users in different markets may have quite different preferences and queries. Chinese users are in a different market segment from users in the US and Western countries, and they place more value on Chinese language contents and related services. For example, Chinese users may search for free MP3 music download, some local movie stars or historical figures, and the Western users don't. To analyze the competition between Google and Baidu in Chinese search market, we divide the indexed web pages into two categories: web pages with Chinese language contents and web pages with Non-Chinese language contents, as denoted by S and S', respectively. M_a denotes the web pages with Chinese language contents indexed by Google, M_{h} denotes the web pages with Chinese language contents indexed by Baidu, M'_{o} refers to the web pages with Non-Chinese language contents indexed by Google, M'_{h} refers to the web pages with Non-Chinese language contents indexed by Baidu. As we have described earlier, Baidu knows more about the Chinese users and offers more Chinese language contents and related services to the Chinese users compared to Google, so $|M_b| >> |M_c|$. Since Google has a strong position in the US and Western countries, $|M'_{o}| > |M'_{b}|$. Here we also define p, the probability that a Chinese user searches for Chinese language contents. 1 - p will be the probability that a Chinese user searches for Non-Chinese language contents. p will be close to one since most of the times a Chinese user searches for web pages with Chinese language contents.



Figure 6 Google vs. Baidu

We define:

- $NB_{g,c}$: net benefit of a Chinese search user if search through Google's search engine
- $NB_{b,c}$: net benefit of a Chinese search user if search through Baidu's search engine
- $Q_{\rm g,c}$: quality of Google's search platform for Chinese users
- Q_{hc} : quality of Baidu's search platform for Chinese users
- $R_{g,nc}$:relevance factor of Google's search platform for web pages of Non-Chinese language contents
- $R_{g,c}$: relevance factor of Google's search platform for web pages of Chinese language contents
- $R_{b,nc}$:relevance factor of Baidu's search platform for web pages of Non-Chinese language contents
- $R_{b,c}$: relevance factor of Baidu's search platform for web pages of Chinese language contents

Since Baidu's search platform is more tailored to Chinese local users, we have $Q_{b,c} > Q_{g,c}$. Since Google has better general search technology and higher relevance factor, so $R_{g,nc} > R_{b,nc}$, but for web pages with Chinese language contents, Baidu has an advantage since it knows better about Chinese language, characters, and culture, so $R_{g,c} \le R_{b,c}$.

Then we have

$$\begin{split} NB_{g,c} &= (1 - \gamma) \alpha Q_{g,c} + \gamma \beta ((1 - p) R_{g,nc} |M'_g| + p R_{g,c} |M_g|) \\ NB_{b,c} &= (1 - \gamma) \alpha Q_{g,c} + \gamma \beta ((1 - p) R_{b,nc} |M'_b| + p R_{b,c} |M_b|) \end{split}$$

p and is close to 1, so we have

$$\begin{split} NB_{b,c} - NB_{g,c} &\cong \gamma \beta p(R_{b,c}|M_b| - R_{g,c}|M_g|) \\ &\approx \gamma \beta(R_{b,c}|M_b| - R_{g,c}|M_g|) \\ &> \gamma \beta R_{b,c}(|M_b| - |M_g|) \\ &> 0 \end{split}$$

Similarly we can have

$$\frac{NB_{b,c}}{NB_{g,c}} \cong \frac{R_{b,c}|M_b|}{Rg,c|M_g|} > 1$$

So we can see that Baidu's search engine platform offers much greater net benefit to Chinese local users, compared to Google's search engine platform. This is why Baidu won over Google and became the leading search engine platform in China.

Similarly, we can derive that Google's search engine platform offers much greater net benefit to Western country users, or those who have needs similar to users in Western counties. This is why Google is so successfully in the U.S. and Western countries.

Even the Chinese local users are not homogeneous. Different segments of Chinese users have different preferences. Google is the preferred search engine among highly educated Chinese, particularly those who have studied in Western countries (Wang, 2010). This observation complies with our model, in which Google offers larger net benefit to Western users and those who have similar needs as Western users

Google's Failure in Korea

Google started its Korean site in 2000, much earlier than the time when it entered Chinese search market (Herman, 2006). In 2010, ten years after Google's entrance to Korean search market, Google still did not have a strong foothold in Korea. Google was reported to only capture 4% of the search market in Korea (Bonfils, 2010). The top three are Naver, Daum, and Nate, all local companies in Korea. They have market shares of 62%, 21%, and 10%, respectively (Bonfils, 2010).

Naver was founded in 1999, and became the leading search engine in 2003 (Chun, 2007). It has been able to strengthen its lead position since then, and has become the dominant search platform in Korea. Naver's success is often attributed to its popular knowledge sharing service, which was launched in 2002. The knowledge sharing service allows Naver users to post questions on any subject to its site, and select the best answers from those provided by other Naver users. In 2007, 44,000 questions were posted by Naver users and 110,000 answers were received each day, making Naver the world's fifth largest search portal (Choe, 2007). Naver's knowledge sharing service was copied by other Internet companies and has become the must-have feature for Internet portals in Korea. The Internet portals keep the

questions and answers in their private databases not shared with other portals or with search engines like Google. When a user conducts a web search, the search engine yields relevant items from its own Q&A database along with traditional search results from web pages on the Internet (Choe, 2007). By July 2007, Naver's usergenerated database had accumulated 70 million entries (Choe, 2007).

Why has Google, the most powerful search engine on this planet with the most advanced technology, huge financial resources and well-known brand name, only captured a mere 4% of search market share in Korea ten years after its entrance? Google's failure in Korea has some similarities and differences compared to its fate in China. As in China, Google competes in a market with quite different language, culture in Korea compared to the US and Western countries. One example is the design of the home page of the search site. Google's bare bone design of home page with a single search box is very popular in Western countries, but in Korea, people like web pages with rich contents. Korean users prefer Naver's way of homepage design, a page with detailed category listings, animated pictures, online shopping and new headlines so users can find a lot of information on one page easily (Herman, 2006). Google tried to impose the same minimal design of web page to Korean users but it failed to do so. Recently, almost ten years after Google's entrance to Korea, Google Korea changed its home page design to be more like other Korean local web sites, to cater for the preferences of Korean users (QUT, 2010).

One key difference of Korea's search market compared with China and Western countries is the lack of Korean language contents (Herman, 2006). When Google entered Korea, the number of web sites in Korean language was relatively small, compared to web sites in English and Chinese. With this unique characteristic, no matter how powerful Google's search engine is, there are not enough Korean language contents for Google to search on to satisfy Korean users. Google's strength is to find the answer when the content is rich, but this is not the case in Korea. Google's advantage in advanced search technology was greatly reduced. Naver's knowledge sharing service helped it to combat the lack of Korean language contents on the web. By leveraging the user-generated contents, Naver can better satisfy the users who search for information on the Internet.

Next we will apply the two-sided markets model to analyze the competition between Naver and Google Korea to show that Naver offers greater net benefit to Korean users compared to Google Korea. In previous sections, we always assumed that all search engines have equal access to all the web sites on the Internet. In the competition between Google and Naver in Korea, it is not the case. Besides the publicly accessible web sites in Korean language, there is some proprietary database that is owned by one company and is not shared with others. As in Naver's case, Naver has its own knowledge sharing database that holds large number of questions and answers and it does not share with other companies. We model the competition between Google and Naver as a two-sided market competition in Figure 7. M_p denotes the pages in Korean language that are publicly accessible, M_c denotes the contents stored in Naver's private database that only Naver's users have access to. In the early stage, since there were not enough web sites in Korean language, M_p was pretty small. M_c was initially zero, but it has grown very fast and today M_c has a lot more contents than M_p . Google only has access to M_p , Naver has access to both M_p and M_c .



We define

- NB_{g} : the net benefit of a Korean user if use Google's search platform
- NB_n : the net benefit of a Korean user if use Naver's search platform
- ${\it Q}_{\rm g}~$: quality of Google Korea's search platform
- Q_n : quality of Naver's search platform
- R_g : relevance factor of Google Korea's search platform on Korean contents
- R_n : relevance factor of Naver's search platform on Korean contents
- R'_{n} : relevance factor of Naver's search platform on its own private database

So we have

$$\begin{split} NB_{g} &= (1 - \gamma) \alpha Q_{g} + \gamma \beta R_{g} |M_{p}| \\ NB_{n} &= (1 - \gamma) \alpha Q_{n} + \gamma \beta (R_{n} |M_{p}| + R'_{n} |M_{c}|) \end{split}$$

Based on our analysis above, Q_n is larger than Q_g since Naver knows more about Korean language and culture and can better cater for the needs of Korean users. R_g has about the same value as R_n because Google has advanced technology in search in general contents but Naver knows more about Korean characters, and contents. γ is large due to the strong positive cross network effect. At the early stage of the competition, $|M_p|$ was small due to lack of Korean language contents on the Internet, $|M_c|$ was close to zero since the service was just launched and there were not much contents.

So we have

$$NB_{g} = (1 - \gamma)\alpha Q_{g} + \gamma \beta R_{g} |M_{p}|$$
$$NB_{n} = (1 - \gamma)\alpha Q_{n} + \gamma \beta R_{n} |M_{n}|$$

 NB_g would have similar value as NB_n , meaning that Naver's understanding of Korean users, language, and culture offsets Google's technology advantage in search. With the launch of the knowledge sharing service by Naver, we have observed exponential growth of this service and $|M_c|$ has become larger and larger. $|M_p|$ also grew since more web pages in Korean language became available, but at a much slower pace, because knowledge sharing service has been so popular in Korea. By July 2007, Naver's user-generated database had accumulated 70 million entries (Choe, 2007). Even though we could not accurately estimate the value of $|M_p|$ and $|M_c|$, it is reasonable to estimate that $|M_c|$ quickly surpassed $|M_p|$ and has widened the gap since then, thus we have $|M_c| >> |M_p|$. So we have $NB_n >> NB_g$. That means Naver can offer much greater net benefit to Korean users compared to Google Korea after it has accumulated a lot of questions and answers in its private knowledge sharing database. This explains Google's failure in Korea.

Another observation about Google Korea is that it not only failed to be the market leader, but it was also left far behind. Google Korea is the No. 4 search platform in Korea. Besides Naver, Daum, and Nate, the No. 2 and 3 search platforms, also offer knowledge sharing service and have their own private knowledge sharing databases, though not as large as Naver's. So Daum and Nate can achieve a similar competitive advantage over Google Korea with their knowledge sharing services and thus they have greater market shares.

Proposed Strategies for Search Market Leaders

Based on the mathematical model and our two-sided markets model analysis for the search engine market competition in previous sections, we propose the following strategies for Google and local market leaders.

Strategies for local market leaders

- Don't copy Google's way. Understand local customers and make the site tailored to them.
- Offer other services that local customers value and that can increase the stickiness.
- Work with local partner web sites to extend customer base beyond searchers coming to its own site and include those searchers coming to its partner web sites.
- Offer a service that can build a private database that is only accessible from its own site. For example, add a knowledge sharing service.
 - If it already has such a service and is open, make it closed.
- Leverage existing customer base and build new two-sided markets platform with strong positive cross network effect.

Strategies for Google

- Continue to maintain and strengthen leadership in search technology.
- Catch up in search technology in other languages such as Chinese and Korean.
- Understand customers in each market and offer service that is tailored to the needs of local customers in each market.
- Work with local partner sites to extend customer base beyond searchers coming to its own site and include those coming to its partner web sites.
- In a local market that lacks contents in local language, build a knowledge sharing service for people to get answers to their questions.
- Offer service that can build a private database that is only accessible from its own site. For example, knowledge sharing service.
- Leverage its dominant position in the US and Western countries to better serve local merchants/businesses.

- For example, some Chinese and Korean local companies may want to expand to markets where Google is a search market leader, so Google has an advantage in helping them compared to Baidu and Naver.
- Leverage existing customer base and build new two-sided markets platform with strong positive cross network effect. For example, Google Android, Google Checkout, Google AdSense.

Conclusion and Discussion

Google has achieved huge success in the US and Western countries but failed to compete in some major Asian countries such as China and Korea. In this paper, we modeled the search engine market as a two-sided market to analyze the search engine market competition, and applied the two-sided markets model to analyze the competition in various scenarios.

First we pointed out that the search engine market is a special type of twosided market. The specialness lies in the fact that all search engines have access to all the publicly accessible web pages, so a search engine can easily overcome the chicken-and-egg problem by indexing a large set of web pages. Given this specialness, a general search engine market competition is similar to the competition in a market with a value chain model.

We then used the two-sided markets model to analyze and explain the market competition in the early stage of the search engine history, why Google could emerge as the market leader, and why Google is so powerful. In the early stages of the search engine history, each search engine only covered a small and different portion of the total web space. Through the model analysis, we showed that searchers could derive more value by using more than one search engines during one search query. This is why multiple search engines could co-exist in the early stage of the search engine history. Google, a late comer in the search market, with advanced search technology, greatly improved the search relevance and thus rose to dominance. Google also worked with partner web sites to build a new two-sided markets platform, AdSense. We pointed out that this strategy would introduce strong positive cross network effect between partner web sites and advertisers and thus help strengthen Google's market lead position. In general, we suggested that a search engine market leader should create new two-sided markets platform with strong positive cross network effect by leveraging its existing customer base.

Next we applied the two-sided markets model to analyze Google's failure in China and Korea. In Chinese local search market, Baidu has an advantage in understanding the Chinese local users and offers services that the Chinese local customers value. Baidu also has advantage in Chinese language. Chinese users are in a different market segment and Google's advantage in general search technology and huge indexed web pages lost value in this market segment. Failure to understand Chinese local users and provide services they value is one of the critical reasons for Google's failure in China.

Google's failure in Korea is different from its experience in China. Compared to the US and China, there were not enough contents in Korean language when Google entered Korean search market. To fill that gap, local search company Naver offered knowledge sharing service, which allows the users to post and answer questions online. The online knowledge sharing service grew rapidly and enabled Naver to provide better search results compared to Google, who only had access to the public web space. In general, when the contents on the Internet are scarce, building a knowledge sharing service and keeping it closed will give a search engine platform an advantage over its competitors.

In the end, we proposed some strategies to Google and local search market leaders on how they could maintain and strengthen their leadership positions in their respective markets. One key strategy is to build a new two-sided markets platform with strong positive cross network effect by leveraging existing customers.

The paper has some limitations. One limitation of our analysis in this paper is that we make quite a few linear assumptions on the functions used in the model. This won't impact the results in this paper since a linear function is a monotonic increasing function. But a more complex function could improve and make it more realistic. For example, we assume that g(M) is a linear function to simplify the analysis. But in reality, the marginal value of M decreases. When M is large, the same amount of increase in M would have much less impact on the value searchers derive compared with the case when M is small.

In this paper, we devised strategies for search engine market leaders. Future potential topics include developing strategies for new entrant search engine companies and discussing how a new entrant search engine company should compete in a search engine market under different market conditions. Another future potential research topic is to examine other countries in which Google does not have leader-ship position, to understand why that happened and derive some insights from that. Did Google fail with similar reasons as in China and Korea, or did something different happen?

References

- Alexa. (2010). *Top sites in United States*. Retrieved from www.alexa.com/topsites/ countries/US.
- Armstrong, M. (2004). Two-sided markets: Economic theory and policy implications. Paper prepared for ENCORE workshop on competition issues in twosided market, Oct. 28, 2004.
- Armstrong, M. (2005). Competition in two-sided markets. *Industrial Organization*, *EconWPA*.
- Barboza, D., & Stone, B. (2010). *China, where U.S. Internet companies often fail.* Retrieved from www.nytimes.com/2010/01/16/technology/16failure.html.
- Beebe, J. (1977). Institutional structure and program choices in television markets. *Quarterly Journal of Economics*, 91(1), 15-37.
- Bonfils, M. (2010). Search around the world: South Korea. Retrieved from searchenginewatch.com/3640458.
- Bradlow, E., & Schmittlein, D. (1999). The little engines that could: Modeling the performance of World Wide Web search engines. *Marketing Science*, 19(1), 43-62.
- Caillaud, B., & Jullien, B. (2001). Chicken & Egg: Competing matchmakers. *Rand Journal of Economics*, *34*(2).
- Chatterjee, P., Hoffman, D., & Novak, T. (1995). Commercial scenarios for the web: Opportunities and challenges. *Journal of Computer Mediated Communication, special issue on Electronics Commerce, 1*(3).
- Choe, S. H. (2007). *South Koreans connect through search engine*. Retrieved from www.nytimes.com/2007/05/technology/05online.html.
- Chou, T. (2011). *Bing vs. Google: Which is better*? Retrieved from www.businessinsider.com/bing-vs-google-which-is-better-2011-3.
- Chun, G. (2007). *Naver Myth of internet age success*. Retrieved from www.koreaittimes.com/story/3083/naver-myth-internet-age-success.
- Csorba, G., & Hahn, J.-H. (2003). Functional degradation and two-sided network effects: An application to software markets. *JEL*.
- Daltorio, T. (2010). *The number one reason why Google failed in China... and it's a big one*. Retrieved from www.investmentu.com/2010/January/google-fails-in-china.html.
- Finin, T. (2009). BlindSearch evaluates Google, Bing and Yahoo search engines. Retrieved from ebiquity.umbc.edu/blogger/2009/06/07/blindsearch-evaluatesgoogle-bing-and-yahoo-search-engines/.
- Gandal, N. (2001). The dynamics of competition in the Internet search engine market. *International Journal of Industrial Organization*, 19(7), 1103–1117.

- Google. (2010). *Google quarterly report* Retrieved from biz.yahoo.com/e/101029/ goog10-q.html.
- Hammonds, H. K. (2003). How Google grows... and grows... and grows, fast company? Retrieved from www.fastcompany.com/magazine/69/google. html?page=0%2C0.
- Herman, B. (2006). *Google fails to make inroads in South Korea* Retrieved from www.usatoday.com/tech/news/2006-04-30-google-south-korea_x.htm.
- Hitwise. (n.d.). Search engine market share. Retrieved from www.hitwise.com/us.
- Hoffman, D., & Novak, T. (1996). A new paradigm for electronics commerce. *The Information Society*.
- Jansen, B., Zhang, M., & Zhang, Y. (2007). The effect of brand awareness on the evaluation of search engine results. Computer Human Interaction 2007 Conference.
- KARMA SNACK. (2011). Search engine market share Retrieved from www.karmasnack.com/about/search-engine-market-share/.
- Kopp, C. (2011). Weekly web watch: Bigger better Bing? Retrieved from www.minyanville.com/businessmarkets/articles/google-search-microsoft-bing-yahoosearch/2/15/2011/id/32815.
- Kunder, D. M. (2010). The size of the World Wide Web. Retrieved from www.worldwidewebsize.com/.
- Lawrence, S., & Giles, L. (1998). Searching the World Wide Web. *Science*, 280(3), 98-100.
- Lawrence, S., & Giles, L. (1999). Accessibility of information on the web. *Nature*. 400.
- Liebowitz, S. J., & Margolis, S. E. (1994). Network externality: An uncommon tragedy. *Journal of Economic Perspectives*, 8(2), 133-150.
- MarketTheGlobe. (2010). *Search engine leader by country* Retrieved from www. marktheglobe.com/SEO-Usa.
- McGee, M. (2010). *Bing hits all-time high market share: Nielsen* Retrieved from searchengineland.com/bing-hits-all-time-high-market-share-nielsen-38280.
- Mukhopadhyay, T., Rajan, U., & Telang, R. (2002). *Competition between Internet search engines*. Hawaii International Conference on System Sciences.
- Netcraft (2010). October 2010 web server survey. Retrieved from news.netcraft. com/archives/2010/10/12/october-2010-web-server-survey.html.
- NETMARKETSHARE. (2010). Search engine market share. Retrieved from marketshare.hitslink.com/search-engine-market-share.aspx?qprid=4.

- Parr, B. (2009). *Google vs. Bing: The blind taste test.* Retrieved from mashable. com/2009/06/07/blindsearch/.
- Pogue, D. (2009). *Bing, the imitator, often goes Google one better*. Retrieved from www.nytimes.com/2009/07/09/technology/personaltech/09pogue.html.
- QUT. (2010). *Why has Google failed in Korea*? Retrieved from asiancorrespondent. com/31699/why-has-google-failed-in-korea/.
- Rappa, M. (2009). Case study: Google, managing the digital enterprise. Retrieved from digitalenterprise.org/cases/google.html.
- Rochet, J., & Tirole, J. (2003). Platform competition in two-sided markets. *Journal* of the European Economic Association, 1(4), 990-1029.
- Rochet, J., & Tirole, J. (2004). Two-sided markets: An overview. *Rand Journal of Economics*.
- Rochet, J., & Tirole, J. (2006). Two-sided markets: A progress report. *The RAND Journal of Economics*, 37(3), 645-667.
- Roson, R. (2005). Two-sided markets: A tentative survey. *Review for Network Economics*, *4*(2).
- Rysman, M. (2000). Competition between networks: A study of the market for yellow pages. *Review of Economic Studies*.
- Salibra, L. (2010). Google's China lessons: Forbidden fruit & face saving retreat. Retrieved from larrysalibra.com/2010/01/13/googles-china-lessons-forbiddenfruit-face-saving-retreat/.
- SearchEngineWatch. (n.d.). Google's market share in the past. Retrieved from askville.amazon.com/Google%27s-market-share-2006-2005-2004-2003-2002/ AnswerViewer.do?requestId=5588947.
- Steiner, P. (1952). Program patterns and preferences, and the workability of competition in radio broadcasting. *Quarterly Journal of Economics*, *66*(2), 194-223.
- Sun, M. C., & Tse, E. (2007). Sustainable growth of payment card networks: A twosided market approach. *Journal of Business Strategies*, 24(2).
- Telang, R., Mukhopadhyay, T., & Wilcox, R. (2000). An empirical analysis of Internet search engine choice. Working Paper, Carnegie Mellon University, Pittsburgh, PA.
- Telang, R., & Mukhopadhyay, T. (2005). Drivers of web portal use. *Electronic Commerce Research and Applications*, 4, 49-65.
- Telang, R., Rajan, U., & Mukhopadhyay, T. (2004). The market structure for Internet search engines. *Journal of Management Information Systems*, 21(2), 137-160.
- VentureBlog. (2010). 4 keys to Google's success. Retrieved from www.ventureblog. com/2003/05/4-keys-to-googles-success.html.

- VisualEconomics. (2010). Planet Google: From philosophies to market shares. Retrieved from www.visualeconomics.com/2010-02-03-planet-google-fromphilosophies-to-market-shares/.
- W3C. (2000). *A little history of the World Wide Web*. Retrieved from www.w3.org/ History.html.
- Wang, T. (2010). My suggestions for making Google's services more relevant for non-elite Chinese users. Retrieved from culturalbytes.com/post/340498962/ googleandchina.
- White, A. (2008). Search engines: Left side quality versus right side profits. *Working Paper, Toulouse School of Economics.*
- Zhao, W. G., & Tse, E. (2011). *Advantages of the two-sided markets model*. Manuscript submitted for publication.

Biographical Sketch of Authors

Wugang Zhao is a senior engineering manager at Paypal Inc.. He received his Ph.D. from Stanford University. His research interests include entrepreneurship, innovation, competitive analysis, two-sided markets.

Edison Tse (Ph.D., Massachusetts Institute of Technology) is an Associate Professor of Management Science and Engineering, Stanford University. He received the Donald Eckman Award from the American Automatic Control Council. He had served as an Associate Editor of the IEEE Transactions of Automatic Control, and a co-editor of the *Journal of Economic Dynamics and Control*, which he co-founded.

He designed and directed a Stanford Executive Certificate Program in China, a Stanford Policy Analysis Certificate Program for Chinese government officials, and a Stanford Financial Engineering Certificate Program in Hong Kong. His current research interest is in new business model innovation and how it is applied to business and industry transformation.