Contents Creation by Users and the Sustainability of Community Based Search on the Web

Seonyoung Shim[a] and Byungtae Lee[b]

Graduate School of Management, Korea Advanced Institute of Science and Technology, Chengryangridong, Dogdaemungu, Seoul, Korea
Tel: +82-2-958-3656, Fax: +82-2-958-3604, Email: [a]syshim@kgsm.kaist.ac.kr, [b]btlee@kgsm.kaist.ac.kr

Abstract

The role of the search portal is evolving from a gateway to a destination by the advent of the community-based search. A community-based search delivers content from the knowledge community inside the portal that binds the traffic to the portal instead of allowing it access to external information sites. The Q&A mechanism available to users gives more relevant search results, hence resolving the search limitations inherent in the automated search engine. However, by monopolizing traffic inside the portal, it may of course deprive external sites. We examine how the evolution of the community-based search may affect information ecology on the web from a long-term perspective. The results show that the community-based search can be an efficient service if there is an increase in total web quality when the magnitude of network externality is high, there are sufficient references to external sites before knowledge sharing in the community, and the marginal disutility of users due to irrelevant information is low. However, in the opposite environment, this service causes the collapse of the web by lowering the quality of the content for both portal and external sites.

Keywords
Community-Based Search, Knowledge Community

Introduction

The search engine has evolved from a simple engine to a giant portal site which has attracted much attention from researchers. Dewan et al. describe the natural formation of portals, their co-existence with satellites, and the asymmetry in the quantity of content and advertisements in portal [7,8].

In the turbulent cut-throat competitive environment of the internet, the early winners of the U.S. search portal market, such as WebCrawler, Lycos, Infoseek, Excite and Hotbot, faded away. Next, Yahoo started a directory service but, users were not satisfied with this and wanted more comprehensive search results which could keep pace with the fast growth of the web. AltaVista's search results included additional information, called "links." Links revitalized the web. By analyzing the links, Google revolutionized it further by providing search results in a logical order based on users' click choice [6].

Recently portals have taken advantage of User-Created-Content(UCC) provided by internal knowledge communities, incorporating it into their search results and calling it 'community-based searching.' In a knowledge community, one person can submit a question, and anybody who has a response shares his/her ideas by answering the question, drawing from both explicit and tacit knowledge. Thus accumulated UCC becomes an important information source for future inquiries. Moreover, the ask-and-answer mechanism can provide more direct and personalized solutions to users' information needs. Because community-based searching is gaining popularity, search portals are trying to introduce this service [4].

The merit of incorporating this new service into web directory services is well manifested by the surprising success of Naver, the first in Korea to offer UCC. It became the almost instant winner in the search market as soon as UCC was introduced in 2002. Chae and Lee's research [4] on Naver identifies the success factors of the community-based search from four perspectives: Relevance, Engagement, Convenience, and Co-Creation. Even though Naver is a laggard among search portals in Korea, it could defeat global giants such as Yahoo and Google based on these factors within the knowledge community. Because of the continued success of community-based searching, other portals are being forced to embrace this trend. Yahoo and Google also started the 'Yahoo.answer' and 'Google.answer' service after they observe the great success of Naver. We are interested in changes in traffic patterns on the web because of this new service. Figure 1 shows that knowledge-related, community-based searches drastically increased the total traffic flow into Naver between July 2002 and July 2003.

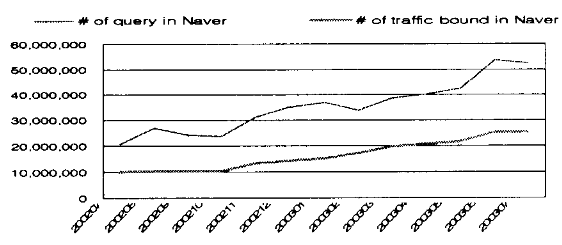

Figure 1 - Queries and amount of bound traffic in Naver

At the same time, it also boosted the amount of bound traffic within the portal. Roughly half of all inquiries are satisfied either by the portal's internal content accumulated from previous community-based searches, or instantly provided answers by other users for new inquiries.

While traditional directory services direct traffic to other linked web sites, UCC confines it within portals. Hence, search portals are being transformed from gateways into destinations in terms of finding information. If this trend continues, portals will accumulate more and more content, and may eventually monopolize web traffic.

Portals have served as the hubs of the web's structure, and many "satellite" sites have thrived on the flow of traffic directed by those hubs. What happens to those satellites if confinement of traffic within portals continues to increase? This is the main focus of this paper. It may seem obvious that satellites will eventually be starved to death because of portals' monopolizing traffic. However, further investigation of the knowledge transfer between portals and satellites reveals a serious and interesting question for research. Much UCC within portals is actually synthesized and customized from the content of satellites. Answering users' questions requires not only a responder's own personal knowledge, but also that of other sites that he/she knows better than the inquiring person. In other words, content in satellite sites is the main source of shared information within a portal's knowledge community.

Of course, traffic flow is the ultimate incentive to any site. Hence, reduced traffic by empowered portals will lower the motivation to invest in content updates for those satellite sites. Will this then finally cause the collapse of the information ecology which the Internet has so successfully built up? Should portals therefore deliberately consider this tradeoff between community-based searching and long-term profitability and sustainability? In order to answer these questions, we investigate the conditions under which information ecology collapses or thrives by means of an analytical model.

Model

We analyze the Internet environment consisting of one search portal and n satellite sites. The portal offers both the community-based search and a general directory service. Each satellite site provides information content at no charge. In our model we assume that all information users select a search portal as the entry point for web searching. After searching on the portal site, users select the most relevant site link for further information. However, if their questions are answered by one of the community-based search results, they will remain within the portal.

Game of Attention

Portals and satellites try to lure more customers by providing valuable content, which in turn increases their advertising revenue. Higher quality content attracts more users, but requires more investment. Moreover, investment in satellite sites has a positive impact on portal content because information from satellites is an important root source for portals, as explained above. There is also a tradeoff in terms of investment strategy from the long-term perspective. If a portal invests more in its own internal content, it benefits from a high level of traffic binding in the current stage. However, the corresponding decline in demand for the satellite sites may considerably lower incentive to invest in satellites, finally resulting in withdrawal of some satellites from the market. Because the collapse of the satellite market may be critical to their content, portals should carefully consider satellite investment, traffic binding, and their effects on long-term sustainability.

To capture these features between portal and satellite investment tradeoffs, we take the two-stage spatial competition model and assume that all the satellites are symmetric. In addition, there is free entry into the market until there is no market surplus; hence the profit of each satellite goes to zero. The subscript j $(j = p, s)$ denotes the player – either portal or satellite site – and i $(i = 1,2)$ denotes the stage of the game.

Cost Functions

Let $I_{j,i}$ and $q_{j,i}$ denote the investment in and quality of the content of the player j (j= p, s) in stage i. For the satellite site, content quality is identical with investment level, which costs $I_{s,i}$: $q_{s,i} = I_{s,i}$ (1)

Portal content consists of information accumulated within the knowledge community inside the portal, which is determined by investment from satellites, the total number of satellites in the market, and its own investment.

$$q_{p,i} = \alpha n_i I_{s,i} + \beta I_{p,i}$$

As each satellite invests more, the quality of portal content increases. Additionally, as the number of satellites increases, the portal has access to more content sources, which can deliver information on various topics. High quality information means first that the information should fit the user requirements, and secondly that the content should be rich. A large number of satellites provides various topics of information and increases the probability of information relevancy. If α% of the content in satellite sites is replicated in the portal's knowledge community, the satellite sites may contribute to the portal's quality as much as $\alpha n_i I_{s,i}$ $(0 < \alpha < 1)$.

In case of portal, they invest in the platform for the community, while the content is produced by the users within the community. However, the quality of the content from portal investment is somewhat ambiguous, while a satellite's investment is straightforward, as shown in Equation (1). Portal content is dependent on the quality of user involvement: the quality of the knowledge shared by community users. This kind of investment can deliver higher returns than direct investment in content when the quality of user involvement is high ($\beta > 1$); otherwise it can derive lower returns ($0 < \beta < 1$).

The investment and quality structure of both players – portal and satellites – implies that a portal's investment is directly related to satellite quality, but that satellite investment can be complementary to portal quality. Thus the relationship between portal and satellites is asymmetric.

Utility and Demand Functions

The portal and its satellites play a spatial competition game based on their content quality. For the analysis of this game, we develop a hybrid spatial model of a linear city in the center of a circle, on the circumference of which n number of symmetric satellites are uniformly distributed, which is the equilibrium result of the circular city model. At the top of the linear city, there is a search portal, and at the bottom, there is a searched satellite, which has the most relevant content for the information searcher. In other words, the search mechanism locates the best satellite site for the users at the end of the linear city from the circle around it, as shown in Figure 2.

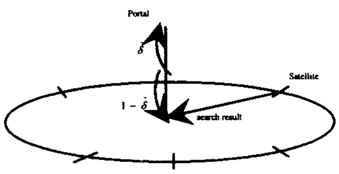

Figure 2 - Hybrid spatial competition model

We assume that all users are uniformly distributed in the linear city depending on their information needs after they search the portal. Because we focus on the users' staying or leaving the portal after searching, which satellite is located at the end of the linear city is not important for deriving the demand function. Users select the site that offers greater utility: either the portal or one of its satellites. The utility function from each site is defined as follows:

$$U_{p,i}(\delta) = (q_{p,i} + \gamma \tilde{\delta}_{i-1} m_{i-1})(1 + \gamma \tilde{\delta}_i m_i) - \lambda\delta$$
$$U_{s,i}(\delta) = n_i q_{s,i} - \lambda(1-\delta)$$

δ_i and m_i denote the ratio of traffic bound inside the portal and the total amount of traffic in stage i ; hence $\tilde{\delta}_i m_i$ means the amount of traffic which finalizes the portal as a destination site. γ denotes the magnitude of network externality in the knowledge community. The more users, the higher the utility, because the content in the community – e.g. Q&A – is enriched by the larger number of users. Moreover, the content improves not only due to the number of users in the current stage, but also due to that from the previous stage, because the content cumulates as time goes on. Therefore, we consider both $\tilde{\delta}_i m_i$ and $\tilde{\delta}_{i-1} m_{i-1}$ in the utility functions for portals. By contrast, for satellites, we assume there is no user involvement in content creation, eliminating the possibility of network externality effects on content quality. However, the total number of satellites, n_i, expands the content from the satellite sites and increases the probability of users moving to satellites, so that we must consider it in the utility function of the satellite site. Therefore a user who finds perfect information within the portal gets a utility of $(q_{p,i} + \gamma\tilde{\delta}_{i-1} m_{i-1})(1+\gamma\tilde{\delta}_i m_i)$ while the user who finds perfect information from a satellite gets as much as $n_i q_{s,i}$.

δ means the location of a user's best relevant information and λ denotes the marginal disutility by unit distance from that information. Thus the user incurs a cost of $\lambda\delta$ when going to the portal and $\lambda(1-\delta)$ when going to a satellite site. As mentioned earlier, much of the content accumulated within portals through knowledge community-based services comes in fact from external sites that participating users have visited. Therefore let α denote the degree of duplication of portal content with that of other satellites. The higher this duplication becomes, the less differentiated they become. Then, $1-\alpha$ implies the differentiation between the two sources. It can be assumed that the magnitude of λ is negatively related to the size of α. Hence, we assume that $\lambda = \mu(1-\alpha)$, where μ is the coefficient. Therefore the utility function can be rephrased such that:

$$U_{p,i}(\delta) = (q_{p,i} + \gamma \tilde{\delta}_{i-1} m_{i-1})(1 + \gamma \tilde{\delta}_i m_i) - \mu(1-\alpha)\delta$$
$$U_{s,i}(\delta) = n_i q_{s,i} - \mu(1-\alpha)(1-\delta)$$

Thus the traffic binding ratio, $\tilde{\delta}_i$, is determined where a user's utility is the same for any site – portal or satellite – and is between 0 and 1.

$$U_{p,i}(\delta) = U_{s,i}(\delta) \quad (2)$$

$$0 < \tilde{\delta}_i < 1 \quad (3)$$

Profit Function

From the demand analysis, $m_i \tilde{\delta}_i$ and $m_i(1-\tilde{\delta}_i)$ represent the demand for the portal and satellites in stage i and, due to the symmetric condition, one satellite gets traffic of $m_i \frac{(1-\tilde{\delta}_i)}{n_i}$. Additionally, we assume that the revenues of portal and satellites are proportional to the number of satellites because each satellite is a potential advertiser for the portal site. When each site has c thousands of CPM, the profit functions of the portal and a satellite are defined as follows under the assumption that all site visitors are exposed to the banner advertisements:

$$\pi_{p,i} = cn_i m_i \tilde{\delta}_i - I_{p,i} \qquad \pi_{s,i} = cn_i m_i \frac{1-\tilde{\delta}_i}{n_i} - I_{s,i}$$

By the assumption of a monopolistically competitive and contestable market for satellites, satellite sites enter the market until there is no expected surplus. Both the portal and satellites have perfect information about this, so they consider it in their profit maximization in each stage. $\bar{n}_i$ denotes the maximum number of satellite sites in stage i:

$$\pi_{s,i}(\bar{n}_i) = 0 \tag{4}$$

Result and Discussion

We set up a two-stage game model of portal and n symmetric satellites and investigated the optimal strategies of both players by backward induction. We also focus on the environmental changes in web knowledge: with traffic bound in the portal, how is the quality of the portal and individual satellite sites affected? How does it change the number of satellites and information ecology on the web? In the first stage, the portal starts the knowledge community service. Part of the traffic is bound inside the portal with a ratio of $\tilde{\delta}_1$, and entries into satellite market occur until $\pi_{s,1}(\bar{n}_1)=0$. In the second stage, the portal continues its service with a traffic binding ratio of $\tilde{\delta}_2$, and the satellite market also satisfies $\pi_{s,2}(\bar{n}_2)=0$. First, we start from the case when the total amount of traffic is unchanging. In the ensuing analysis, we consider a more realistic situation, investigating the effect of increasing traffic with each stage.

Case 1: When the Number of Users is Fixed

Here, we assume that the total amount of traffic is 1 in both stages ($m_i = 1$). It is common knowledge that traffic has been dramatically increasing since the beginning of the Internet age. However, in this section, we assume a fixed amount of traffic. By comparing the results from this case with those from the next case, where traffic is increasing, we can examine in detail the effect of increasing traffic.

In the second stage, a user's utility functions from portal and satellites are defined respectively as follows:

$$U_{p,2}(\delta) = (q_{p,2} + \gamma\tilde{\delta}_1)(1+\gamma\tilde{\delta}_2) - \mu(1-\alpha)\delta \tag{5}$$

$$U_{s,2}(\delta) = n_2 q_{s,2} - \mu(1-\alpha)(1-\delta) \tag{6}$$

By using Equation (2), thus plugging $\tilde{\delta}_2$ to δ into Equations (5) and (6) based on the rational expectation equilibrium, we can derive $\tilde{\delta}_2$, where a user becomes indifferent between the two players:

$$\tilde{\delta}_2 = \frac{-\beta I_{p,2} + (1-\alpha)(n_2 I_{s,2} - \mu) - \gamma\delta_1}{\gamma(\beta I_{p,2} + \alpha n_2 I_{s,2} + \gamma\delta_1) - 2\mu(1-\alpha)} \tag{7}$$

The objective functions of the players are given as follows:

$$\max_{I_{p,2}} \pi_{p,2} = cn_2\tilde{\delta}_2 - I_{p,2} \tag{8}$$

$$\max_{I_{s,2}} \pi_{s,2} = c(1-\tilde{\delta}_2) - I_{s,2} \tag{9}$$

From the FOCs of Equations (8) and (9), the optimal investment levels of the portal and satellite sites are obtained:

$$I^*_{p,2} = \frac{1}{\gamma}\left[\frac{\mu(1-\alpha)\{2-\alpha(2+\gamma)\}}{\beta} - \frac{c\beta n_2}{(\alpha+\beta)^2}\right] - \frac{\gamma\delta_1}{\beta}$$

$$I^*_{s,2} = \frac{1}{\gamma}\left[\frac{\mu(1-\alpha)(2+\gamma)}{n_2} - \frac{c\beta}{(\alpha+\beta)^2}\right]$$

By plugging these into Equation (7), the traffic binding level(δ^*_2) and maximum number of satellites($\bar{n}_2$) is:

$$\delta^*_2 = \frac{1}{\gamma}\left\{\frac{1}{(\alpha+\beta)} - 1\right\} \tag{10}$$

$$\bar{n}_2 = \frac{\mu(1-\alpha)(\alpha+\beta)^2(2+\gamma)}{\{c(\alpha+\beta)^2(\gamma+1) - \alpha\}} \tag{11}$$

In the first stage, users consider their total utility from both stages. Because there is no switching cost of moving from the portal to satellites or vice versa, users' expected total utility when they choose the portal or satellite in the first stage is defined such that:

$$U_{p,tot} = U_{p,1} + E(U_2) = U_{p,1} + \{\tilde{\delta}_2 U_{p,2} + (1-\tilde{\delta}_2)U_{s,2}\} \tag{12}$$

$$U_{s,tot} = U_{s,1} + E(U_2) = U_{s,1} + \{\tilde{\delta}_2 U_{p,2} + (1-\tilde{\delta}_2)U_{s,2}\} \tag{13}$$

The subscript *tot* denotes the sum of results from stages 1 and 2. If $\tilde{\delta}_2$ portion of traffic is bound to the portal at stage 2, the probability of a user's using the portal is also $\tilde{\delta}_2$, so that the expected utility function of the second stage is defined as $\{\tilde{\delta}_2 U_{p,2} + (1-\tilde{\delta}_2)U_{s,2}\}$. Irrespective of the user's choice in stage 1, the expected utility in the second stage is identical; hence the demands for portal and satellite sites in the first stage are determined where $U_{p,1} = U_{s,1}$.

$$U_1^p(\delta) = U_1^s(\delta) \tag{14}$$

$$U_1^p(\delta) = (q_1^p)(1+\gamma\delta_1) - \mu(1-\alpha)\delta \tag{15}$$

$$U_1^s(\delta) = n_1 q_1^s - \lambda(1-\delta) \tag{16}$$

By taking the rational expectation equilibrium again in Equation (14), $\tilde{\delta}_1$ and the objective functions are derived such that:

$$\tilde{\delta}_1 = \frac{-\beta I_{p,1} + (1-\alpha)(n_1 I_{s,1} - \mu)}{\gamma(\beta I_{p,1} + \alpha n_1 I_{s,1}) - 2\mu(1-\alpha)} \qquad (17)$$

$$\max_{I_{p,1}} \pi_{p,tot} = \pi_{p,1} + \pi_{p,2}(\bar{n}_2, I^*_{p,2}, I^*_{s,2}, \delta_1(I_{p,1}, I_{s,1}, n_1))$$

$$= (c\eta\delta_1 - I_{p,1}) + (\tfrac{\gamma}{\beta}\delta_1 + A) \qquad (18P)$$

$$where A = \frac{\mu(1-\alpha)(2+\gamma)\{\beta + (\alpha+\beta)(1-(\alpha+\beta))\}}{\gamma\{(\alpha+\beta)^2(1+\gamma) - \alpha\}} - \frac{\mu(1-\alpha)\{2-\alpha(2+\gamma)\}}{\gamma\beta}$$

$$\max_{I_{s,1}} \pi_{s,tot} = \pi_{s,1} + \pi_{s,2}(\bar{n}_2, I^*_{p,2}, I^*_{s,2}, \delta_1(I_{p,1}, I_{s,1}, n_1))$$

$$= c(1-\delta_1) - I_{s,1} \qquad (18S)$$

For simplicity of analysis, we set the discount rate at 1, and by considering the condition $\pi_{s,2}(\bar{n}_2) = 0$, $(\pi_{p,2}, \pi_{s,2})$ becomes $(\frac{\gamma}{\beta}\delta_1 + A, 0)$. The optimal investment levels at the first stage for the two players respectively are:

$$I^*_{p,1} = \frac{1}{\gamma}\left[\frac{\mu(1-\alpha)\{2-\alpha(2+\gamma)\}}{\beta} - \frac{(c\beta n_1 + \gamma)^2 n_1}{c\beta((\alpha+\beta)n_1 + \frac{\gamma}{c})^2}\right]$$

$$I^*_{s,1} = \frac{1}{\gamma}\left[\frac{\mu(1-\alpha)(2+\gamma)}{n_1} - \frac{(c\beta n_1 + \gamma)n_1}{((\alpha+\beta)n_1 + \frac{\gamma}{c})^2}\right]$$

In the same way, δ_1^* and $\bar{n}_1$ are given:

$$\delta_1^* = \frac{1}{\gamma}\left[\frac{1}{\{(\alpha+\beta) + \frac{\gamma}{c n_1}\}} - 1\right] \qquad (19)$$

$$\bar{n}_1 = \frac{\gamma}{c}\left[L(j,k) + \frac{j^2}{L(j,k)} + j - (\alpha+\beta)\right]^{-1} \qquad (20)$$

where

$$L = \sqrt[3]{k + j^3 + \sqrt{k^2 + 2kj^3}}$$

$$j = \frac{\{c(\alpha+\beta)+\gamma\}(\gamma+1)}{3\mu(2+\gamma)(1-\alpha)}, \quad k = -\frac{\{c(\alpha+\beta)+\gamma\}\alpha}{2\mu(2+\gamma)(1-\alpha)}$$

In the first stage, the portal and satellites should consider the potential result from the second stage: in the second stage, there are so many satellites as to result in the zero profit satellite market. Hence by considering the expected profit under this condition, the portal and satellites try to optimize their total profit in the first stage. δ_1^* and δ_2^* in Equations (19) and (10) show the trend of traffic binding in the portal and satellites.

Proposition 1 *Under the condition of a fixed amount of traffic on the Internet, the community-based search causes the portal to monopolize traffic while the number of satellites diminishes.*

Proof *By Equations (10) and (19), Proposition 1 is supported because* $\delta_1^* < \delta_2^*$ *for any* $\bar{n}_1$.

Proposition 2 *Under the condition of a fixed amount of traffic and a monopolistically competitive satellite market, the portal's increasing monopolization of traffic causes satellites to invest less in its content at each progressive stage, so that the content quality of satellite sites gradually degrades stage by stage.*

Proof *By Equations (4), (9) and (18), it is to be observed that*

$$I^*_{s,1} = c(1-\delta_1^*)$$
$$I^*_{s,2} = c(1-\delta_2^*)$$

Therefore, according to the result of proposition 1, it can be concluded that $I^*_{s,2} < I^*_{s,1}$ (21)

Case 2: When the Number of Users is Growing

Here, we assume a more realistic case when the amount of traffic increases with each period. This type of growth rate for traffic is called Moore's Law on the web because Internet traffic approximately doubles each year. To capture this feature, we let the amount of traffic m_i be i, making the total amount of traffic in the first stage 1, in the second stage 2, and so on: $m_i = i$ (22)

Backward induction shows that only the second stage derives different solutions than were found in case 1.

By Equation (22), δ_2 and the objective functions are given such that:

$$\tilde{\delta}_2 = \frac{-\beta I_{p,2} + (1-\alpha)(n_2 I_{s,2} - \mu) - \gamma\delta_1}{2(\beta I_{p,2} + \alpha n_2 I_{s,2} + \gamma\delta_1)\gamma - 2\mu(1-\alpha)}$$

$$\max_{I_{p,2}} \pi_{p,2} = 2cn_2\tilde{\delta}_2 - I_{p,2} \qquad (23)$$

$$\max_{I_{s,2}} \pi_{s,2} = 2c(1-\tilde{\delta}_2) - I_{s,2} \qquad (24)$$

By the same flow as in case 1, we can derive the solutions of this case such that:

$$I^*_{p,2} = \frac{1}{\gamma}\left[\frac{\mu(1-\alpha)\{1-\alpha(1+\gamma)\}}{\beta} - \frac{c\beta n_2}{(\alpha+\beta)^2}\right] - \frac{\gamma\delta_1}{\beta} \qquad (25)$$

$$I^*_{s,2} = \frac{1}{\gamma}\left[\frac{\mu(1-\alpha)(1+\gamma)}{n_2} - \frac{c\beta}{(\alpha+\beta)^2}\right] \qquad (26)$$

$$\bar{n}_2 = \frac{\mu(1-\alpha)(\alpha+\beta)^2(1+\gamma)}{c\{(\alpha+\beta)^2(2\gamma+1)-\alpha\}} \qquad (27)$$

$$\delta_2^* = \frac{1}{2\gamma}\{\frac{1}{(\alpha+\beta)}-1\} \qquad (28)$$

From the solutions in Equations (25) ~ (28), we know that the profits of both sites in the second stage are identical to the results of case 1. Therefore we skip the analysis of stage 1 because all the solutions are same as in case 1.

$$(\pi_{p,2}(\bar{n}_2),\ \pi_{s,2}(\bar{n}_2)) = ((\frac{\gamma}{\beta}\delta_1 + A),\ 0)$$ where

$$A = \frac{\mu(1-\alpha)(2+\gamma)\{\beta+(\alpha+\beta)(1-(\alpha+\beta))\}}{\gamma\{(\alpha+\beta)^2(1+\gamma)-\alpha\}} - \frac{\mu(1-\alpha)\{2-\alpha(2+\gamma)\}}{\gamma\beta}$$

Proposition 3 *Under the condition of an increasing amount of traffic, the community-based search causes both the portal and satellites to get more traffic in each progressive stage.*

Proof. *By Equations (19) and (28), it is easily shown that the traffic bound in the portal increases in the second stage:*

$$m_1\tilde{\delta}_1 = \frac{1}{\gamma}\{\frac{1}{\{(\alpha+\beta)+\frac{\gamma}{c\bar{n}_1}\}}-1\},\quad m_2\tilde{\delta}_2 = \frac{1}{\gamma}\{\frac{1}{(\alpha+\beta)}-1\}$$

$$m_1\tilde{\delta}_1 < m_2\tilde{\delta}_2$$

Similarly, it is also straightforward that the traffic of each satellite $(m_i(1-\tilde{\delta}_i))$ *also increases in the second stage according to (29).*

$$m_1(1-\tilde{\delta}_1) < m_2(1-\tilde{\delta}_2) \qquad (29)$$

$$\text{because } m_1(1-\tilde{\delta}_1) - m_2(1-\tilde{\delta}_2)$$
$$= (1-\tilde{\delta}_1) - 2(1-\tilde{\delta}_2) = (\tilde{\delta}_2 - 1) - \tilde{\delta}_1 < 0$$

In this case, there is more traffic to both the portal and satellites. In addition, satellites' investment level is also increasing under the conditions $\pi_{s,2}(= m_2c(1-\tilde{\delta}_2) - I_{s,2}) = 0$ and growing demand, which is the opposite result from case 1. In the previous fixed demand case, the effect of decreasing traffic to satellites reduced their investment levels, but the increasing traffic in this case lessens this effect and makes the satellites also feel that they will gain more and more traffic in each progressive stage.

Proposition 4 *Under the condition of an increasing amount of traffic, each satellite increases its investment level in the second stage so that the investment level is larger than when traffic is fixed. However, the number of participants in the satellite market at this stage,* $\bar{n}_2$*, is less than when traffic is fixed.*

Proof. *By Condition (4) and Equations (24) and (18), it is straightforward that*

$$I_{s,i}^* = cm_i(1-\delta_i^*) \qquad (30)$$

Here, $I_{s,1}^*$ *is identical both in case 1 and case 2. Therefore, from Equations (21), (29) and (30), it can be concluded that*

$$I_{s,2}^{*\ case\ 1} < I_{s,1}^{*\ case\ 1,2} < I_{s,2}^{*\ case\ 2}$$

In case of $\bar{n}_2$*, by Equations (11) and (27), it is clear that*

$$\bar{n}_2^{\ case2} < \bar{n}_2^{\ case1}$$

The fixed demand increases only the portal's traffic by reducing individual satellite traffic. However, the growing demand increases the traffic of both sites in the second stage, even though the traffic binding ratio, is decreasing. Hence satellite investment is increasing, but the growing market reduces the number of satellites. It may seem counterintuitive. The increased cost (investment) causes fewer satellites to enter the market. This implies that the growing demand causes fiercer competition between the portal and satellites, and the result is that fewer but more competitive satellites survive in the market compared to case 1.

Proposition 4 shows that $\bar{n}_2$ in case 1 is smaller than in case 2 due to the increasing market cost effect. However, we are mainly interested in the comparison of $\bar{n}$ from each stage in order to examine the performance of the community-based search: How does the number of satellites change the quality of portal content? Does the total web quality improve or degrade? This analysis is not so easy because the solution of $\bar{n}_1$ is quite complex, as shown in Equation (19). Moreover, $q_{p,1}^*$ is also a function of $\bar{n}_1$. Thus we move on numerical analysis in order to examine changes in $\bar{n}$, $q_{p,1}^*$ and total web quality and to identify the efficient and inefficient conditions of the community-based search from the perspective of total web quality.

Numerical Analysis

In this section, we investigate the impact of various parameters on information ecology through numerical analyses. For this purpose we define the total web quality in stage i as follows:

$$q_{t,i} = q_{p,i} + \bar{n}_i q_{s,i}$$

This definition captures not only the depth, but also the extent of the available information by considering the number of satellites. The number of satellites represents the breadth of information, while the quality of satellite sites reflects the depth of information. In the ensuing discussion, satellites content quality is represented as $\bar{n}_i q_{s,i}$.

Information Replication and Synthesis (α)

Figures 3a and 3b show the effect of α on web quality when $\gamma = 1.5, c = 1, \mu = 4$ and $\beta = 0.2$. The bold line shows the results of the first stage. The solid and dotted lines stand for the results of the second stage from case 1 and case 2; the dotted line is the result under growing traffic and the solid one is from fixed traffic. As shown in section 4, the results in the first stage are the same in any case; thus the bold line can be applied to both cases 1 and 2.

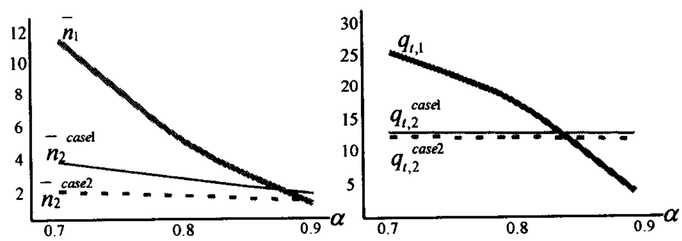

Figure 3a. Changes in portal and satellites quality

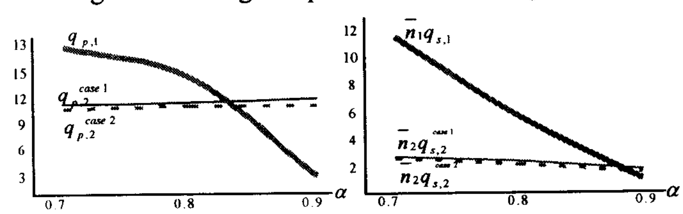

Figure 3b. Changes in $\bar{n}$ and total web quality

Figure 3a demonstrates the change in portal and satellite content quality with increasing α, and Figure 3b shows the changes in $\bar{n}$ and total web quality. By summing the two results in Figure 3 – portal and satellite content quality – we can derive the total web quality, shown at right in Figure 3b ($q_{t,i} = q_{p,i} + \bar{n}_i q_{s,i}$). As the figures show, the maximum number of satellites and content quality is decreasing in α. The decline is more conspicuous in the first stage, where the size of $\bar{n}_1$ and $q_{t,1}$ is drastically reduced. Hence $q_{t,1}$ is higher than $q_{t,2}$ under low α and lower under high α. This means that vigorous knowledge copying from satellites to portal – high α – makes the total web quality (q_t) increase as time goes on.The trend is the same whether the demand is fixed (case 1, solid line) or growing (case 2, dotted line). However, growing demand causes $q_{t,1}$ to cross $q_{t,2}$ at a higher level of α because the value of $q_{t,2}$ is always lower than the fixed traffic case. This results in a larger range of α, which satisfies $q_{t,1} > q_{t,2}$, so that the growing demand increases the possibility of the collapse of information ecology on the web more than when the demand is fixed under low α.

Network Externalities (γ)

Our next analysis is on γ. Our main interest is whether thecommunity-based search has an effect on its quality. In order to focus on the comparison of total web quality in the first and second stages instead of just overall web quality, we adopt a joint analysis of γ and α in Figure 4. In this way, the relative size of $q_{t,1}$ and $q_{t,2}$ can be clearly demonstrated. The bold and solid lines stand for $q_{t,1}$ and $q_{t,2}$ under low γ ($\gamma = 1.0$), while the dotted lines stand for $q_{t,1}$ and $q_{t,2}$ under high γ ($\gamma = 1.5$) when $\lambda = 1.2, \beta = 0.2, c = 1$.

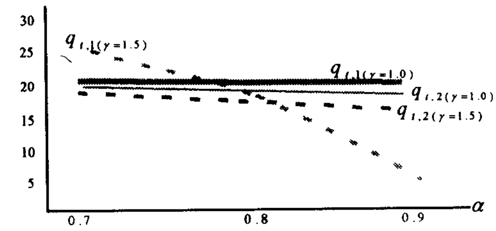

Figure 4. Changes in overall web quality

In Figure 4, the range of $q_{t,1} < q_{t,2}$ is larger under high γ than low γ, which means a large magnitude of network externality (γ) in the knowledge community increases the probability of $q_{t,1} < q_{t,2}$.

Proposition 5 *The community-based search can be an efficient service delivering an increase in total web quality when network externality within the knowledge community is high, and when the content of the portal depends greatly on the satellites' content.*

Proof *Proposition 5 can be clearly proved from the impact of the extent of knowledge replication and synthesis (α) and network externalities (γ). Figures 3 and 4 show that the probability of $q_{t,1} < q_{t,2}$ increases with high α and γ. Under these conditions, $\bar{n}$ and I_p^* increase in stage 2. Although I_s^* decreases in the second stage with fixed user base, the effect of increased $\bar{n}$ dominates the decrease of I_s^*, finally the content quality of the portal, q_p^*, and the content quality of its satellites, $\bar{n}q_s^*$, increase in the second stage.*

Remember that when the user base is growing, the increase of I_s^* was already proved through proposition 4. Therefore the increase in overall web quality is obvious with increased $\bar{n}$, I_p^* and I_s^*. Therefore, in any case under high α and γ, the web gets richer in terms of both portal and satellite quality. When the community-based search provides more accurate and useful information than the content from satellites, it increases the total number of satellites and improves overall web quality as well. In proposition 5, a high level of γ means that as more and more users participate in the knowledge community, the content of the portal site becomes more useful and provides greater benefit to users. A high α also means that users do a certain amount of referring to content from satellites, implying that community users share sufficiently useful information for satisfying their various information needs by actively interacting with satellite sites in order to foster a developing web environment.

Proposition 6 *The community-based search can be an inefficient service resulting in a decrease in overall web quality when network externality within the knowledge community is low and when the content of the portal is not dependent on the satellites' content.*

Proof *The results also clearly prove the impact of the extent of knowledge replication and synthesis (α) and network externality (γ). Figures 3 and 4 again show that the probability of $q_{t,1} > q_{t,2}$ increases with low α and γ. In this case, $\bar{n}$ and I_p^* decrease in the second stage. The effect of a lower $\bar{n}$ induces lower q_p^* and $\bar{n}q_s^*$ in the second stage. Therefore under these conditions, both portal and satellite quality decrease sharply.*

A low level of α and γ imply inefficient behavior on the part of knowledge community users. For example, if they share knowledge among themselves instead of consulting professional sites content quality in the knowledge community is degraded because their information is less valuable. The benefit of sharing knowledge among users is therefore lower, which means a low level of γ.

Conclusion

From this analysis, we can conclude that the size of $\bar{n}$ is a critical determinant of information ecology on the web. Initially, portals were not the independent giants on the web. Despite traffic monopolization, portal content quality greatly depends on the number of satellites, meaning that the health of the portal may be connected to the rise and fall of satellites. A portal's myopic tendency toward traffic monopolization may be helpful for its short-term profitability. However, from the long-term perspective, it can result in the death of both satellites and portal, and ultimately the fundamental collapse of the web under inefficient conditions for the community-based search. Therefore, in the developing web environment, as the community-based search provides more benefit to information users, the role of satellites is extremely important. By contrast, the synergy of information sourcing from professional information sites and the synthesis or refinement of it into customized knowledge for community users can deliver ultimately sustainable benefits to information users and result in the improvement of information ecology on the web.

In this paper, we investigated the effect of the community-based search on the information ecology of the web. As more information becomes available on the web, the need for ways to find relevant information increases due to the complexity of the information search. The community-based search seems to emerge as a winner within search portals by combining the benefit of a search engine and user-created knowledge. However, so-called "user created knowledge" often comes from other web pages. Individuals synthesize information from various external or "satellite" web sites in order to provide more personalized and relevant information. Therefore, binding more traffic inside portals may endanger the web information ecology under certain conditions we have identified. The results of our study emphasize the importance of the co-existence of both portal and satellites and the delicate balancing of traffic and useful content. Portals must understand the positive feedback loop in the web economy and give sufficient consideration to the strategic "coopetition" – cooperation and competition – with satellites for long-term sustainability, both for themselves and for the overall health of the web.

References

[1] Adamic, L.A. and Huberman, B.A. (2000). "The Nature of Markets in the World Wide Web", *Quarterly Journal of Electronic Commerce*, Vol. 1.

[2] Alstyne, M.V. and Brynjolfsson, E. (2005). "Global Village or Cyber-Balkans? Modeling and Measuring the Integration of Electronic Communities", *Management Science*, Vol. 51, pp. 851-868.

[3] Battelle, J. (2005). *The Search: How Google and Its Rivals Rewrote the Rules of Business and Transformed Our Culture.* Portfolio.

[4] Chae, M. and Lee, B. (2005). "Transforming an Online Portal Site into a Playground for Netizen", *Journal of Internet Commerce*, Vol. 4 pp. 95-114.

[5] Coffman, K.G and Odlyzko, A.M. (2000). "Internet Growth: Is there a "Moore's Law" for data traffic?", Working paper, AT&T Lab.

[6] David, V. and Mark, M. The Google Story, Delacorte Press (2005)

[7] Dewan, R.M., Marshall L. F. and Zhang J. (Winter 2002-2003) . "Management and Valuation of Advertisement-Supported Web Sites", *Journal of Management Information Systems*, 19, 3 87-98.

[8] Dewan, R.M., Marshall L. F. and Zhang J. (2002). "Managing web sites for profitability: Balancing content and advertising", *Proceedings of the Thirty-fifth Hawaii International Conference on System Science.*