# A Case Study of the Effects of Platform Software Selection on Information System Maintenance Cost - An Example of Enterprise Search System Establishment -

Keiko Shimazu
Keio University
4-1-1 Hiyoshi, Kohoku-ku, Yokohama,
Kanagawa 223-8521, Japan
shimazu@z6.keio.jp

Souhei Morita
Keio University
2-17-22 Mita, Minato-ku, Tokyo
108-0073, Japan
ok.souhey@dmc.keio.ac.jp

Kaoru Mori
Keio University
2-17-22 Mita, Minato-ku, Tokyo
108-0073, Japan
kaoru@mdbl.sfc.keio.ac.jp

Yusuke Okumura
Keio University
2-17-22 Mita, Minato-ku, Tokyo
108-0073, Japan
chai@sfc.wide.ad.jp



**Abstract.** Enterprise search systems have been focused on since around 2005. However, there are very few case-study reports/papers how to build enterprise search systems and how to provide their superior cost performance. We propose an importance of maintenance cost for system management, especially after system launch. Search engines, well known as Internet search, are generally employed as middleware of enterprise search systems. We tried an experimental study of estimating the system maintenance cost, in order to validate our assumption; the maintenance cost of enterprise search systems is based on an installation cost of those middleware search engines whose main characteristic is said to be simplicity of installation. We employed multiple search engines and performed experiments. We acquired results that the maintenance cost required for the installation of search engines. The results confirmed that our assumption is true to some extent.

## Introduction

The establishment of enterprise search service using search engine software has been gaining more attention since around 2005 (Mukherjee and Mao 2004) (Hawking 2004) (Dmitriev et al. 2006) (Freund and Toms 2006). Enterprise systems were originally information systems established to improve the efficiency of operations by merging business processes over constraints of organizational structure or physical distance. The enterprise system which is most commonly known is ERP (Enterprise Resource Planning system). The role of ERP is to optimize management resources within a corporation across departments. On the other hand, it is the role of Enterprise Search to search databases over constraints of organization structure and physical distance in an integrative manner. Enterprise search differs from data-warehouses in that it enhances decentralization for storing and releasing information, and integrates information whenever necessary. This realizes the minimization of system management cost since each department or person in charge manages the content (excluding essential information). Simultaneously this realizes integration of different sorts of information, with the quality nearly equal to when information is controlled centrally. Introduction of search engines for enterprise search systems as the foundation of information share and utilization, coupled with the movement of internal control, is gaining momentum. By providing services which not only cross borders of organization, but also pick up information from outside

corporations, companies have a high possibility of realizing a breakthrough to the global share of information; this issue will be described in more detail in (3) of the next chapter titled "Enterprise Search Systems".

On the other hand, when information systems are launched and begun to be used by end-users, the focus of the system management generally transfers to the following three points.

(a) operation: this is divided into operations for end-users to use the provided functions and operations for engineers to activate, stop and adjust the systems.

(b) maintenance: a task for maintaining the projected functions and performance. This generally indicates tasks which are conducted daily by systems engineers while considering the balance of the entire system and avoiding unexpected effects on other parts. This includes careful engineering manipulation operation and provision of supplies. This does not have to include tasks which require skills or expertise. It points to all the tasks which are required for maintaining the system in good condition.

(c) enhancement: when operating a system, the functions or performance which were initially assumed may not realize their expected effectiveness only through maintenance tasks. The enhancement is achieved by adding new functions or service modules or by developing the successor.

The following sentences show further consideration of the system management costs after initiation of information systems from the perspectives of maintenance and enhancement.

Systems (not only regarding information) generally affect the environment that they are in, and the environment transforms accordingly. The transformed environment creates new requirements for the systems within it. If new requirements don't appear, underlying requirements may surface. Taking these requirements into consideration, a revised system will be developed. The system evolves from the repetition of this process. In this process, the identification of the newly formed requirements and the reflection of these requirements onto the design are conducted in a top-down manner, while the environment change due to the emergence of a new system occurs in a bottom-up manner. Thus, the process of system evolution is a cyclic one, with top-down and bottom-up elements occurring alternately. In this cyclic process, in the case of mechanical systems the revised system is generally launched as successive versions of a product. Three-dimensional CAD systems used for hardware designs are a good example.

On the contrary, for some information systems, successive versions are provided without the suspension of the current systems. Services (such as web services) are representative examples. We can enjoy new features using just web browsers. This difference directly affects the gap between cost management approaches for system management in the two systems. A project which aims to develop successive versions of a mechanical system is generally treated as a different project from the one that developed the previous version. Therefore a new cost management methodology is formed. As for information systems, the successive versions of a system are provided without stopping the previous service, by placing the revised software on the server providing the service. This maintenance sometimes requires minimal cost and low technology. Demands of system customization are managed as issues which should be processed within the enhancement cost of the previous system. When the platform architecture requires change, the revision is managed and proceeds under a new project in most cases. Some stakeholders, especially sponsors, in information system projects hope for the revision to be processed within the system maintenance cost of the previous system.

In other words, in most information system projects, stakeholders expect a platform design where functions and performance can be gradually developed; sponsors urged to other stakeholders to include enhancement cost in maintenance cost. One of the most common methods of realizing this is the function modularization of the minimal units of the system

platform. The other is the preparation of external interfaces of those modules in order to customize and enhance them. These common methods allow various alterations, but also can lead to unexpected behavior, causing frequent problems. There might be a rise in the system management cost. In corporations, this trend is more obvious in information systems than core business systems (ex. accounting systems). Core business systems mount mechanical system management processes, formerly conducted manually, onto computers. If these processes disappeared, the corporate activities would falter. However, in information systems, processes that lead to generation of effects are not clear. These systems find a solution in trial-and-error process, in order to perform the system management with minimum time and effort, and also to multilaterally store and acquire information. Information systems therefore do not have a constant effect on business. They also would not inflict critical damage on organizations even if they did not exist (although this would be inconvenient). Therefore, sponsors (executives in most companies) hope to minimize the system management cost for information systems. Enterprise search systems (information systems, not core business systems) are probably the most notable example. Currently, there are many search engines which are used as the platform for enterprise search systems. There are, however, scarcely any debates or reports on the system management cost of these engines. In this study, we selected the most representative search engines, and actually established systems using them. We then attempted to operate the systems for 6 months, and compared only the system maintenance costs.

This paper is organized as follows: The second chapter titled "Enterprise Search Systems" views the characteristics of enterprise search systems. The third chapter titled “Experiments: Specification” indicates specification of our experiments, and the fourth chapter titled “Experiments : Results” then reports its results. While discussions are described in the fifth chapter titled “Discussion”, the sixth chapter titled “Related Works” introduces related works. The seventh chapter titled “Conclusion” concludes this paper.

## Enterprise Search Systems

**Search Engine.** We believe that system management cost, especially system maintenance cost, differs completely among enterprise search systems, and is significantly dependent on the platform search engines.

Search engines, which are known for their ability to search the Internet, are generally composed as in Figure 1(Hawking 2006). In other words, they consist of 2 databases and 6 process modules. The outline of functions for each process is as follows.

Gathering: collects information on designated websites on the network, or websites linked from these sites, and stores the results in "Collection."

Extracting: for each information source, extracts direction words, date of release, Snippet information, etc. from information stored in "Collection".

Indexing: extracts the relationship between inserted search keywords and information source, and stores it into "Indexes."

Querying: identifies the relevant information source from "Indexes" based on entered search keywords.

Ranking: calculates the significance of the information sources identified by Querying.

Presentation: creates an output based on the calculation performed by Ranking.

**Functions Unique to Enterprise Search.** Why aren't Internet search engines placed on corporation information networks and operated in an as-is manner? This is due to the difference between the requirements of Internet search and enterprise search. (Hawking 2004) and (Lewis 2007) show the results of research on enterprise search from an IEEE journal paper and ACM. There are 4 functions or services which are needed for enterprise search but aren't implemented on Internet search websites. All of them are listed in both (Hawking 2004) and

(Lewis 2007). More importantly, these unique functions of enterprise search enhance or improve the process modules forming the search engine functions, mentioned in the previous section titled “Search Engine”, to fit the environment (each organization) in which the system will be used.

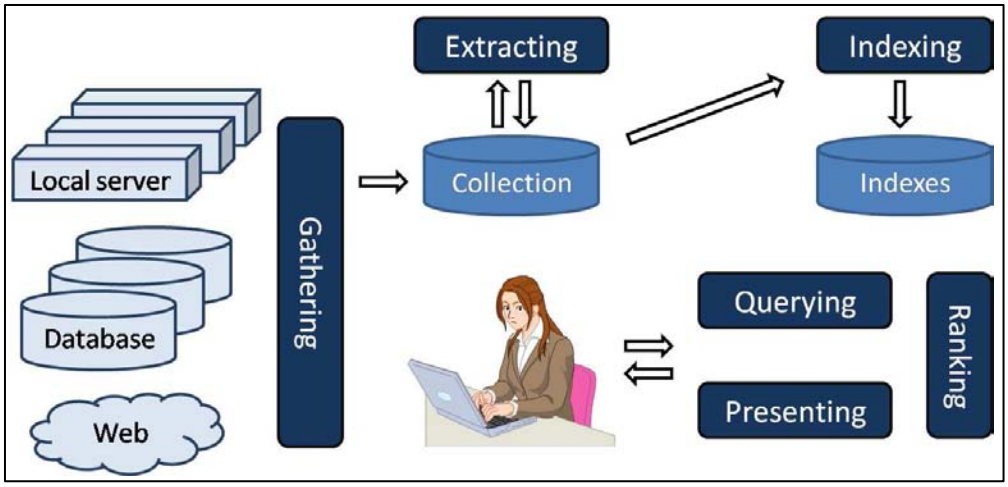


Figure 1. De-facto standard architecture of search engine (Hawking 2006).

**Functions Unique to Enterprise Search.** Why aren't Internet search engines placed on corporation information networks and operated in an as-is manner? This is due to the difference between the requirements of Internet search and enterprise search. (Hawking 2004) and (Lewis 2007) show the results of research on enterprise search from an IEEE journal paper and ACM. There are 4 functions or services which are needed for enterprise search but aren't implemented on Internet search websites. All of them are listed in both (Hawking 2004) and (Lewis 2007). More importantly, these unique functions of enterprise search enhance or improve the process modules forming the search engine functions, mentioned in the previous section titled “Search Engine”, to fit the environment (each organization) in which the system will be used.

(1) Information search for both Shallow Web and Deep Web

Generally, information presented by methods implemented under http protocol, such as html, is called Shallow Web. Information installed on a certain database system architecture is called Deep Web; this database should have a communication interface using http protocols. Search engines on the Internet basically search within Shallow Web. In enterprises, it is required to thoroughly search for information on a corporate network. Modules and crawlers for collecting information need to gather information from Deep Web. There is a need to develop an interface that can access different database system architectures scattered within a corporation. By using this interface, crawlers can collect data within database systems, allowing search engines to search for Deep Web information. This is achieved by enhancing or customizing the process of "gathering" mentioned in the previous section titled “Search Engine”.

(2) Reflecting the importance of unique corporate information

Results of information search on the Internet are calculated and ranked in order of importance by each search engine. This ranking is usually in order of usefulness. However, in an enterprise environment, the ranking algorithm may not be satisfied. Scales reflecting original management strategies, business plans, etc. need to be adopted to sort the results of information search. Moreover, according to the market environment, the importance of each piece of information changes daily. Enterprise search needs to reflect this importance to rank the information and present the results. To achieve this, adjustment of search results and an original calculation method of information importance are required. This is achieved by

enhancing or customizing the processes of "Extracting," "Indexing" and "Ranking" mentioned in the previous section titled “Search Engine”.

(3) Integration of internal and external information search

There is the challenge of realizing management which can respond with agility to the drastically changing market environment. However, the fact that information in databases inside corporations is not enough to solve problems, and that the valuable information which fills this gap exists in websites of the competitors or in CGM (Customer Generator Media, which is created by the general public) makes enterprise search equip integrator results from intranet search and Internet search. This is achieved by enhancing or customizing the processes of "Gathering," "Querying" and "Presenting" mentioned in the previous section titled “Search Engine”.

(4) User Interface

Generally, the specifications of enterprise systems are decided in a top-down manner. To minimize the introduction cost, information systems within corporations or organizations are designed so that end users can understand how to operate them at a glance. This cuts costs for creating manuals or lectures on the operation and maintenance methodology. On the other hand, search engine middleware or COTS (commercial-off-the-shelf) for enterprise search do not assume a specific corporate system. When establishing them as enterprise search systems, there is a need to customize them so that they become a user interface integrated with other systems in individual corporations or organizations. This is achieved by enhancing or customizing the process of "Presenting" mentioned in the previous section titled “Search Engine”. For these reasons, most organizations install search engines on their corporate information web in an as-is manner, and enhance and customize them, in order to develop original enterprise search systems.

## Experiments : Specification

**Policies of Our Experiments.** To make the comparison of maintenance cost of enterprise search systems credible, we (1) measured the actual maintenance cost, under actual operation, rather than calculating the estimate, (2) operated systems around the same time to minimize environmental changes, (3) selected search-engines as platforms of systems for our experimental studies in order to generate broad distinctions among maintenance costs, and (4) assigned different responsibilities to our members and rotated them with the maintenance of each part of systems. These four points will be discussed further in following sections titled “Our Enterprise Search System” to “Search Engines as Our Experimental Materials”. The aim of this verification is to validate the intuitive assumption that maintenance costs of enterprise search systems depend on the search engine selected as the platform. We believe that the maintenance cost of enterprise systems will vary significantly based on the selection of platform search engine software. In addition, we believe that engineers who have used software for a long time would be able to intuitively measure the usability of software. The purpose of our experimental study was to validate the first assumption, while the selection of search engines was based on the second assumption. In other words, we specified the search engines that would be used in this validation based on the second assumption.

More concretely, our approach to the experimental study is as follows:

We actually measured the maintenance costs and compared them, instead of calculating estimates. Specifically, we established enterprise search systems using different search engines and had end-users use them. We attempted to respond to the revision and enhancement requirements during their usage and measured the maintenance cost. However, in case studies of actual systems, in general, the results of experimental studies may take a long time to obtain, and be influenced by issues unrelated to the target attributes. In order to solve the former issue,

we selected two search engines. At the time of selection, we used our experience with installation cost to estimate maintenance cost. Our method of estimation measures the ease of software installation or reinstallation. We first look at how easy the steps on the manual are. Additionally, in order to solve the latter issue, we have limited the area of requirements to ones related to the characteristics of enterprise systems, as shown in a section titled “Functions Unique to Enterprise Search” in the second chapter titled “Enterprise Search Systems”. This is to prevent the comparison of maintenance costs from losing its meaning due to the difference in quality or difficulty of engineering operation required for maintenance and improvement.

**Our Enterprise Search System.** We specified functional specification and performance specification of our enterprise search system, which is actually an intra-university search system, according to the prototypical architecture as mentioned in the second chapter titled "Enterprise Search Systems". We launched two types of our enterprise search system to be actually used by end-users for 6 months each. We selected search engine products which are designed with de facto standardization, as mentioned in a section titled “Search Engine” in the second chapter titled “Enterprise Search System”. These engines in our search systems target the contents inside the university networks and those related to them, and search for ones which correspond to the assigned keyword.

Our systems target contents inside databases in architectures other than HTTP; therefore they provide those contents to the users by issuing queries and using APIs in order to meet the function described in (1) of a section titled “Functions Unique to Enterprise Search” in a chapter titled “Enterprise Search Systems”. These APIs were employed to integrate a faculty profile database, and allow end-users to search data in this database as well as HTML documents.

We customized a ranking algorithm to be utilized for an indication sequence of search results, in order to meet the function described in (2) of a section titled “Functions Unique to Enterprise Search” in a chapter titled “Enterprise Search Systems”. This customization was employed particularly to draw away capital (authorized seasonal) information within search results. This information was released by university head quarters and identified using domain names. The search result on Figure 2 shows an example of a search result using the keyword of "AWARD" (actually "受賞", which means award in Japanese.) We enhanced performance of the search engines to achieve integration search results from intranet and internet, in such a way to implement an interface to Google book search, using its API, in order to meet the function described in (3) of a section titled “Functions Unique to Enterprise Search” in a chapter titled “Enterprise Search Systems”. Our interface was generated based on the API, in order to search items containing keywords related to our enterprise (in this case, our university). Therefore, end-users can acquire search results from intranet and the Google book search site at one time (Figure 3). We customized user interfaces operated by end users, in order to present consistency, as a subsystem of enterprise systems, in order to meet the function described in (4) of a section titled “Functions Unique to Enterprise Search” in a chapter titled “Enterprise Search Systems”. Figure 4 shows the actual design of the top page of our enterprise search systems.

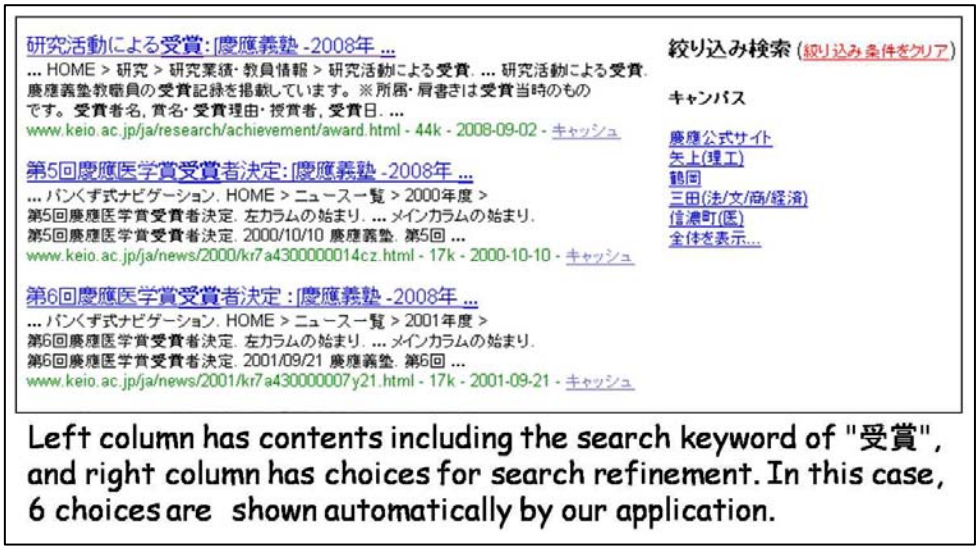


Figure 2. Actual search result of our system : search from intra-university network.


Figure 3. Actual search result of our system : integrated intranet search & Internet search.


Figure 4. Our enterprise search system : `http://kics.dmc.keio.ac.jp/`

**Our Validation Time.** We set the measurement period of our experimental study to 6 months for each enterprise search system. Core business systems were assumed to be manipulated according to the operation processes. Manuals and lessons were prepared, so that the operator could thoroughly understand the operation processes. On the other hand, each individual uses information systems differently. Therefore the end-users devise the utilization methodology on their own, and the systems (similar to search services) will eventually be used daily. In other words, operation efficiency with the introduction of core business systems will improve in a short amount of time (Shimazu, Momma, and Furukawa 03), while the improvement of intellectual productivity with the introduction of information systems would probably require a relatively longer period of time. We then decided to set the period of usage to 6 months in order to measure the maintenance cost of enterprise search systems. It will require several years to adopt a number of search engines, establish an enterprise system for each search engine and maintain them. As mentioned in the first chapter titled “Introduction”, when a system is placed in a certain environment, the environment changes. If several years were required for the experiment, there's a high possibility that the environment surrounding the first system and the last would differ completely. According to our observations during the last few years, the newest IT is spreading to our daily lives faster and faster. If we spent a long time on our experimental study, platform search engines would be improved during this time and web services on the Internet would either change or improve. This would lead to an inaccurate comparison. We have limited this experiment to 2 enterprise search systems, each with a different search engine, and measured the maintenance cost of each system. In selecting the 2 search engines, we considered the fact that our aim is to check for differences between the malignance costs of each enterprise search system. To achieve this, we sought the combination with the largest difference.

We selected the software products with the lowest and the highest cost, established an enterprise search system using each of the products as the platform, operated each system, and took a record whenever maintenance cost was required. We developed two enterprise search systems. We operated both of them on Keio University's intra-university network for 6 months each.

**Search Engines as Our Experimental Materials.** "Nikkei Computer," a computer-related magazine published in Japan by Nikkei Shinbun Inc., featured articles on enterprise search in 2005. In this feature, search engines produced for enterprise search systems are categorized into 6 classes according to their functions and performance (Table 1). To be concrete, there are two sizes of index database capacity of search engines (Y axis in table1), and three search engine characteristics (X axis in table 1). We employed three search engine types from the

table. These are categorized search engines for over 100,000 documents. This is due to the fact that engines which target less than 100,000 documents are more often used as groupware tools (in some cases, in an as-is method) rather than as middleware in enterprise systems. The following describes the characteristics of each engine.

Type 1: In table 1, search engines for enterprise system establishment N1 and N2 are assigned. High-level natural language processing ability, such as creating related keywords from input and processing sentences as input, is the major characteristic in this type.

Type 2: In table 1, search engines for enterprise system establishment I1 and I2 are assigned. When seeds (top URL in the search target) are already specified, the person in charge (without knowledge of the products) can install the system in one day.

Type 3: In table 1, search engines for enterprise system establishment G1 and G2 are assigned. An additional function, which shares information such as what other users used as search keywords, and the usefulness of search results, is the major characteristic in this type.

We selected 2 search engines from each type. Therefore, 6 were chosen for enterprise search system establishment, and each one of them was installed. The goal of the installation process was to make the system operable as an enterprise search system. The installation was conducted using all 6 search engines. For each installation process, we measured the cost (person-hours) required until the system was in an identical state ready for operation.

Table 1: Search Engine types for enterprise search system

| | Type 1 : With advanced natural language processing features | Type 2 : With high cost performance of installation | Type 3 : With advanced groupware features |
|---|---|---|---|
| Upper 100,000 docs | N1<br>N2 | I1<br>I2 | G1<br>G2 |
| Under 100,000 docs | *Unsuitable for enterprise system* | | |

Table 2: Actual installation cost of search engines

| Search engines for enterprise systems | N1 | N2 | I1 | I2 | G1 | G2 |
|---|---|---|---|---|---|---|
| Installation costs (person-hours) | 36 | 29 | 6 | 11 | 16 | 19 |

## Experiments : Results

As described in a section titled “Search Engines as Our Experimental Materials” in a chapter titled “Experiments : Specification”, we adopted the product with the highest and the lowest installation cost as the middleware for establishing two enterprise search systems. We employed search engines as middleware with the highest (N1) and the lowest (I1) installation

cost (described in a section titled “Search Engines as Our Experimental Materials” in a chapter titled “Experiments : Specification”). Using these products, we established two enterprise search systems based on the de-facto standard architecture mentioned in the second chapter titled "Enterprise Search Systems", and had users use them for six months after each systems release on Keio University's. We've measured the maintenance cost required in each system and categorized the results as follows. The former (N1 platform) was operated from October 2006 to March 2007 (Shimazu, Saito, and Yoshinaga 2007), and the latter (I1 platform) from March 2008 to August 2008 (Keio University). Customization of control screen, operations on index database, merging other enterprise systems other enterprise systems using API with, editing of algorithm which decides the rank of the output results, and interface creation operation to allow the database within an enterprise to be a target of search operation. Table 3 shows the average time for a single maintenance operation, calculated from the total required maintenance time. In conducting these five system maintenance operations, we followed the methodology shown in a section titled “Policies of Our Experiments”. 5 engineers formed a rotation and were in charge of the two systems (rotation was formed so that an engineer didn't handle the same category in the two systems). For all the maintenance operations, the enterprise search system with the platform I1, which required the least installation cost, required the least maintenance cost.

Table 3: Maintenance cost

| | N1 | I1 |
|---|---|---|
| Customization of User-interface (average person-hours / event) | 9.2 | 6.0 |
| Updating Index table (average person-hours / event) | 1.0 | 0.4 |
| Plug-in of software or APIs (average person-hours / 1 module) | 17.3 | 2.0 |
| Changing of search result order (average person-hours / 1 change) | 2.5 | 1.2 |
| Interface to databases for gathering their contents (average person-hours / 1 database) | 6.3 | 2.8 |

## Discussion

As described in a section titled “Policies of Our Experiments”, we conducted the installation process (assuming usage is begun by end users) using 6 kinds of search engines. Among the search engines which are middleware for enterprise search system establishment, the cost required for their installation, the simplicity of which is the main characteristic of search engines, seemed to be relatively low compared to others. Furthermore, in the actual installation process of the search engines in case I1 and I2, no engineers from the product company came to help, and the software was sent by email (hardware was also sent in the case of I1). Therefore all the processes were conducted by the authors of this paper. On the contrary, in the case of N1, it was recommended that technicians from the product provider conduct all the setup operations (not for free). In this experiment, the authors, not technicians from product providers, conducted all the operations and recorded the required person-hours in order to compare with other search engine products. Thus we believe that, in some cases (lack of familiarity with search engines), the difference in the person-hours required for installation

may be larger. Of the six search engines, we used the one with the highest (N1) and the lowest (I1) installation cost and established enterprise (Intra-University) search systems with identical functions and performance. After operating each system for 6 months, in terms of maintenance cost, the cost of I1 was far below the cost of N1 in terms of all 5 perspectives mentioned in the fourth chapter titled “Experiments : Results”. If the service period were continued, operation processes would accumulate indefinitely. The graph in Figure 5 shows the total maintenance cost from the 6 months of operation. If the operation period were longer, the difference in cost would be larger.

These validation perspectives are what (Hawking 2004) and (Lewis 2007) list as functions required in enterprise search systems. Therefore our experiment will be informative for the actual operation of enterprise search systems. In the previous chapter titled “Experiments : Results”, we focused on the maintenance cost and presented the results of comparison (table 3). In the actual establishment process, there was a difference in the development cost when the platform was set as N1 and when it was set as I1. Figure 6 shows the total cost, including developing and maintenance cost, displayed in a manner compliant with EVM (Earned Value Management) (Putz et al. 2007).

EVM is one of the methods of managing the progress of project activities such as IT system establishment. The base of EVM is the procurement management tool C/SCSC（Cost/Schedule Control System Criteria）, designed by the U.S. Department of Defense in 1967. In 1998, ANSI presented this as EVMS（Earned Value Management System） with consideration for the private sector. This is the standard of the current EVM. As for government supply, there is a strong movement toward spreading the idea of EVM within the federal government. For instance, the Office of Management and Budget in April 2005 released a note directed towards the CIOs in each of the agencies within the federal government, which urged the introduction of EVM. In December of the same year, the Federal CIO Council released a policy framework related to EVM, supporting each organization of the federal government to create policies of EVM. Along with these movements, many projects set conditions requiring expertise with EVM. Also in Japan, "Conference on Government Procurement of Information Systems" conducted by the Ministry of Internal Affairs and Communications, the Ministry of Economy, Trade and Industry and the Ministry of Finance in March 2000, the report on "the Revision of Government Procurement of Information Systems" was accepted. In this report, EVM was mentioned as one of the ways to make planning and procurement of information systems reasonable. In 2004, ahead of other ministries, the Ministry of Economy, Trade and Industry introduced the progress report by EVM as one of its bidding conditions. Progress management using EVM is also included in "Guideline for optimization of work and systems" (in Japanese), a guideline set at the Government Ministry CIO Conference in March 2006. One of the characteristics of EVM is that not only the progress of activities but also the cost occurrence can be converted into indexes, and managed in the same graph. Utilizing EVM has two purposes. One is to manage the progress of system development performed by third party venders. The other is to enable stakeholders to realize better communication about the progress of their situation. Effects such as discovering the severity of project delay can be expected; therefore stakeholders can decide to overshoot costs of their projects.

On the graph of EVM, when I1 is used as the middleware, the enterprise search system was established with the same cost as initially planned. On the other hand, when N1 was the middleware, development required 7 times as much cost as the initial plan. The development period was 4 months in the case of I1 and 13 months when using N1 (the members' expertise being nearly the same). This is presumably the major reason for the difference in the development cost.

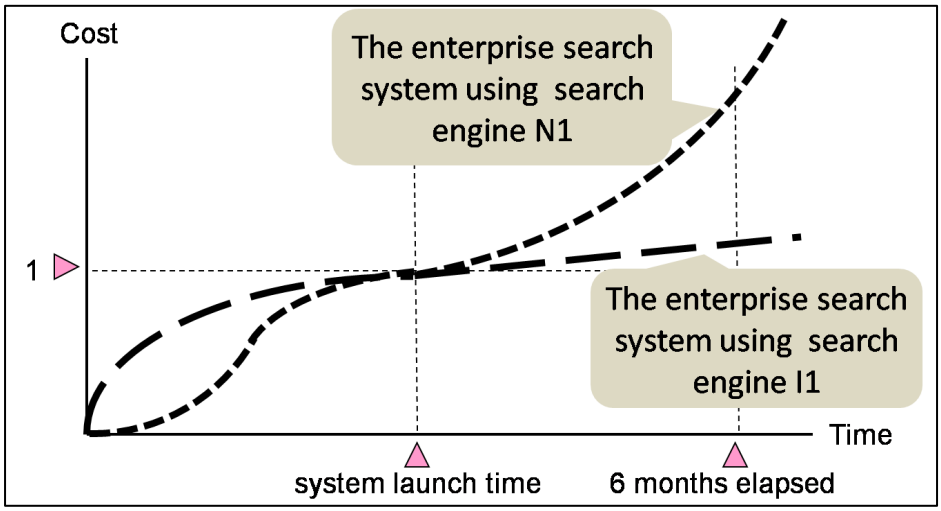


Figure 5. Maintenance cost.

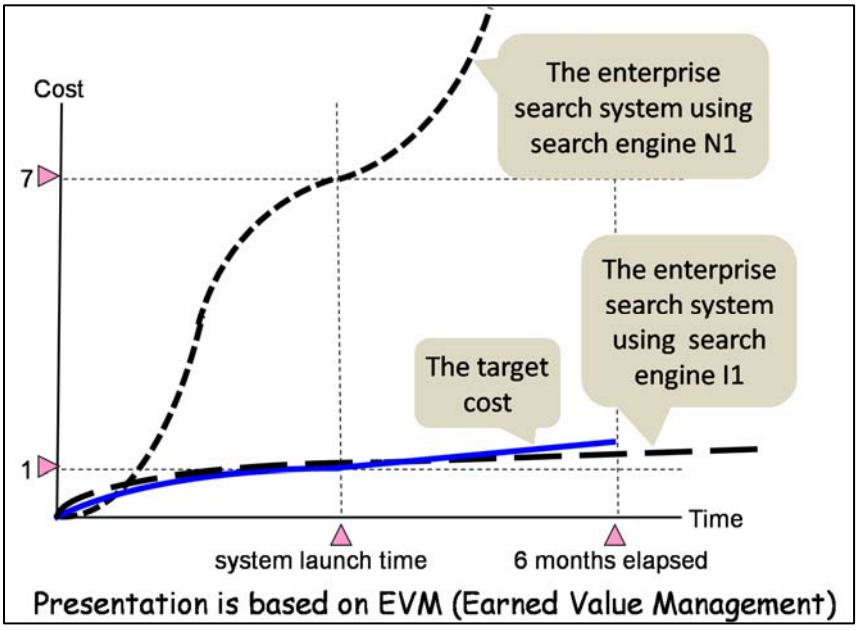


Figure 6. Total cost of building & maintenance.

From these experiments, we have indicated that the facility of initial installation, one of the characteristics of middleware, may possibly affect the development cost of information systems which use these products as platforms.

## Related Works

There are hardly any reports or research papers on middleware (search engine) selection considering both the establishment and the operation of enterprise search systems. There are, however, a number of reports discussing the selection methodology of a wide range of COTS products.

Research related to middleware and COTS selection standards and methodology in information system development began in the 1990's. At the same time, the methodology of information system development employing middleware and COTS became common (Briand 1998). There was a report on how COTS selection methodology developed over the years

(Abdallah, Guenther, and Armin 2007). New methods for COTS selection have also been proposed almost yearly. (Abdallah, Guenther, and Armin 2007) selects 18 of these methods and analyzes them from 7 perspectives. According to this work, (1) most of the COTS selection methods set the requirements and constraints from the stakeholders (including the end-users) as the essential conditions and select the optimal COTS product by evaluating the fitness according to the most important requirements. (2) 2 of them can be applied to multiple COTS products, while others are only effective on certain types of COTS products. (3) One of them determines how incompatible the selected COTS product is with certain objectives. (4) 2 of them allow experts to control each domain. (5) 5 of them are available as tools. (6) Half of them iterate the previous fundamental method of evaluation motioned in (1). (7) One of them compares its compatibility with the other COTS that will be included, and 3 of them focus on the most important requirements and select the one with the most compatibility.

There are many works concerning general COTS selection method. There are, however, scarcely any works reporting on enterprise search.

As can be understood from these facts, our report would be a significant suggestion to the establishment of enterprise search systems, which are now attracting attention from the industrial world.

## Conclusion

In this report, we have focused on enterprise search systems and considered the introduction standards for search engines that would be installed as infrastructure software. We decided to focus on maintenance cost throughout the system lifecycle, since this system would be operated in a spiral-up style. Since there is no general methodology to measure the maintenance cost beforehand, we regarded this as a problem and devised a way to estimate maintenance cost from installation cost. We introduced multiple search engines and performed experiments. We examined our assumption, that we can estimate system maintenance costs based on middleware installation cost. The results confirmed that this assumption is true to some extent. Our claim here lacks scientific evidence. Therefore we have not reached a framework with general versatility. On the other hand, we have witnessed many projects that were shut down because of increasing maintenance cost. Enterprise search systems are no exception. Generally at time of planning enterprise search systems, the focus is only on high performance of search engines. Using the trial validation results reported in this paper as our motivation, we would like to research on engineering which focuses on the safety, stability and ROI of system management in the future.

## References


Abdallah, M, R. Guenther, and E. Armin. 2007. COTS Selection: Past, Present, and Future. *IEEE International Conference on the Engineering of Computer-Based Systems*, 103–114.

Briand, L.C. 1998. COTS evaluation and selection. *Proceedings of the 14th IEEE International Conference on Software Maintenance*, 222-223.

Dmitriev, P.A, N. Eiron, M. Fontoura, and E Shekita. 2006. Using annotations in enterprise search. *Proceedings of the 15th international conference on World Wide Web*, 811–817.

Freund, L, and E.G. Toms. 2006. Enterprise search behaviour of software engineers. *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, 645-646.

Hawking, D. 2004. Challenges in enterprise search. *Proceedings of the 15th Australasian database conference* 27: 15-24.


———. 2006. Web search engines. Part 1. *Computer* 39 (6): 86–88.

Keio University. Keio Information Concierge Service. `http://kics.dmc.keio.ac.jp`. (accessed in November 22, 2008).

Lewis, B. 2007. Guest Editor's Introduction: A Glimpse at the Future of Enterprise Search. *IT Professional* 9 (1): 12–13.

Mukherjee, R, And J. Mao. 2004. Enterprise Search: Tough Stuff. *Queue* 2 (2): 36–46.

Putz, P, D.A. Maluf, D.G. Bell, M.M. Gurram, J. Hsu, H.N. Patel, and K.J. Swanson. 2007. Earned Value Management at NASA: An Integrated, Lightweight Solution. *IEEE Aerospace Conference*, 1-8.

Shimazu, K, A. Momma, and K. Furukawa. 2003. Experimental Study of Discovering Essential Information from Customer Inquiry. *Proceedings of the 9th ACM SIGMOD International Conference on Knowledge Discovery and Data Mining*, 741–746.

Shimazu, K, I. Saito, and S. Yoshinaga. 2007. Content-Compilation System based on User's Contexts *The 21st Annual Conference of the Japanese Society for Artificial Intelligence* (in Japanese).

## BIOGRAPHY

**Keiko Shimazu.** 2005-Present Associate professor of Research Institute for Digital Media and Content, Keio University. 2000 Completed Ph.D in Media and Governance, Keio University. 2000 Senior Researcher of IT Media laboratory of Fuji Xerox Co,. Ltd. 1995 Completed from master course of Graduate School of Keio University, Media and Governance. 1982 Fuji Xerox Co,. Ltd.

**Souhei Morita**. 2008 Researcher of Research Institute for Digital Media and Content, Keio University. 2005 Uchida Yoko co,. Ltd. 2004 Completed from master course of Graduate School of Keio University, Media and Governance. 2002 Admitted master course of Graduate School of Keio University, Media and Governance. 2002 Graduated from Keio University Faculty of Environmental Information. 1998 Admitted to the Keio University Faculty of Environmental Information.

**Kaoru Mori.** 2007-Present Research Assistant of Research Institute for Digital Media and Content, Keio University. 2007 Completed from master course of Graduate School of Keio University, Media and Governance. 2005 Admitted master course of Graduate School of Keio University, Media and Governance. 2005 Graduated from Keio University Faculty of Environmental Information. 2001 Admitted to Keio University Faculty of Environmental Information.

**Yusuke Okumura.** 2007-Present Research Assistant of Research Institute for Digital Media and Content, Keio University. 2007 Admitted master course of Graduate School of Keio University, Media and Governance. 2007 Graduated from Keio University Faculty of Environmental Information. 2003 Admitted to Keio University Faculty of Environmental Information.