

Utilization of Google's Custom Search in Knowledge Finding on Digital Library

Gonang May Perdananugraha

Researcher, Indonesian Institute of Sciences, admin@mail.bit.lipi.go.id

Abstract— Utilization of Google custom search in digital library has gave another option in knowledge finder/tools to explore the information among hundreds, thousands, even millions or more collections that the digital library have. The collections which digital library have are in multi formats including text collections and multimedia collections, and those multi formats collections need to be identified, indexed, and searchable by users. Google custom search provide tools as knowledge finder in searching those collections in web based digital library. Google has known for its amazing search across the world wide web that laid on computer servers around the world, with its proven technology Google offer its custom search engine to be utilized and optimized freely to search and grab certain websites including web based digital library. Using Google custom search also have several weakness among them are un-independent application and of course depends on the internet, the strength using this technology is ability to find information and provide the pattern of information as Google search engine do.

Index Term— Google custom search, knowledge finder, digital library

I. INTRODUCTION

The role of digital library for now days is undoubtable, many beliefs that the present of internet based digital library will replace physical library [1]. Where members or even attenders no need to visit the library, they just have to open the web based application which is library provided for searching the catalogues or even knowledge digging in library collections. Depends on the library itself, some of libraries they provide open systems where users/members can be easily viewing or downloading the documents as they need, or some of libraries just provide the catalogues for users/members to be viewed and searched if they need the full documents further contacts to libraries is required [2][3]. The present of digital library is to help physical library to reach broader scope of their visitors/members, and to boost their collections into knowledge when users optimize the use of that collections.

Many librarians belief that the present of digital library don't be judged as the death of physical library [4]. This is what librarians have to fight with this wrong concept/perception. The physical library should be existed and majestic with the advanced collections and up to date, nice decorations, and great atmosphere, because library as the symbol of knowledge, civilized, and sign of advancement in a nation. while the libraries also have to broaden the services not just in the form of physical but also others services like electronic catalogues. The used of internet also cannot be avoided, where these days many people especially

young generations using internet for their purposes in gaining information and knowledge instead of come to libraries. The growth of the internet should be not seen as threat to libraries, but should be viewed as a chance to use that technologies in boosting their collections and reaping users/members from all around the world[5][6].

One of the important applications in digital library is the tools that used to reap/gather information that collected in digital library. This kind of tools called knowledge mining. Which gathered, analyzed, and find the pattern based on users need. It's kind to be useless if digital library applications don't have such tools, because it will make users hard to find the hidden information among collection of tremendous data/information. This can lead to users dissapointment and unwilling to visit the physical libraries. Compare it if the library have excellent services including in digital library that has comprehensive knowledge mining tools to grab and show it to users, although not all document in open access mode but at least it can help users to know that the information that was looked for are lying in library collections [7].

Many tools in knowledge mining for digital library that have been developed using standard methods like semantic web, indexing based on pages, and words query analysis [8]. No matter techniques that used to build knowledge mining tools, the basic must be based on that three methods. Since the present of popular search engines tools like Yahoo, Google, MSN, and others people tend to rely on these tools to find information/data across the internet, because with these tools that have advanced methods and techniques to know and index million or even billion the web pages that existed in the internet. The interesting question is can we use these advanced technologies that provided by search engines to be used by digital library as knowledge mining, to index, search and show the results based on user needs. At the beginning was really hard to adapt search engine technologies into digital library, to search only in its collections. But along with the advancement of these technologies, for now search engines provider widely open their doors for people in the world can also develop and customize their own needs to use search engines technologies. Especially for Google as search engine provider that full support of open source, open document, and cross platform among computers in the world. They're opening API (Applications Programming Interfaces) for developer around the round to customize any services that Google's provided [9].

By using API's technologies developer or software engineer can easily customizing and attaching Google's search engine technologies into digital library systems, as a knowledge mining tools. The utilizing of Goolge's API

technology doesn't mean that we can take Google's application and put it in digital library as independent application. If we developed Google's API we're still attach to Google's product that will always connected to Google's technology. If we modified Google's search engine into tools for knowledge mining in digital library, we can say we're still using Google's search engine that treat in special way this what we called as Google's customsearch [10].

II. METHODOLOGY

One thing that have to be mentioned is digital library systems that will use Google custom search have to be known by Google search engine. In other words the digital library has to be well developed first and running consistantly in the internet, so Google can crawl and index the digital library web pages to be known and searched via internet [11]. If the digital library just developed and run no more than 1 year, probably the digital library not known very well by Google search engine, or the worst is not known at all by Google. In this case we have to boost digital library by putting some SEO (search engine optimizer programs) into the systems, and make sure the systems has well structure programs and run smoothly in the internet. Also the important once is the server or web hosting that we use in storing digital library must have stabil and reliable hardware and network. If hardware or network is not stabil that will cause the systems also will not run stabil, and this will impact on Google indexing on that digital library systems.

How to test the digital library is well known by Google search engine, we can use Google at www.google.com and search for keywords that indicating the collections or information that available on digital library if one of the result was directing to our digital library, that means our digital library is well known by Google search engine and we can use Google custom search to become tools of knowledge mining in our digital library systems. What if the digital library is in the form of closing systems, that means the collections or information that contained in digital library only can be access by its members. In this way that digital library can't use Google custom search, because Google will not index the collections inside the closing systems. Will be interesting if the digital library is not in fully closing systems but in half closing systems, this means that the collections is on open systems in providing meta documents but digital library close its full documents only by authorized persons. In this way Google search engine still can index its meta documents and direct the result to digital library pages [12].

If the design of digital library in half closed systems, it will attract members/users to get more in touch to digital library systems or even visiting physical library to have full documents, if policy of digital library is requiring to. This will increase the traffic of digital library systems and physical

library.

III. RESULT AND DISCUSSION

After the digital library systems is well known by Google search engine, the next step is utilizing Google custom search into digital library systems to become knowledge mining tools. Before dealing with Google's API (Applications Programming Interfaces) to put API's code programs into digital library, we have to have Google custom search account by registering our selves into this website address www.google.com/cse/. If you already have Google Mail account then you already have Google custom search account either. Like shown in this Figure 1 below:

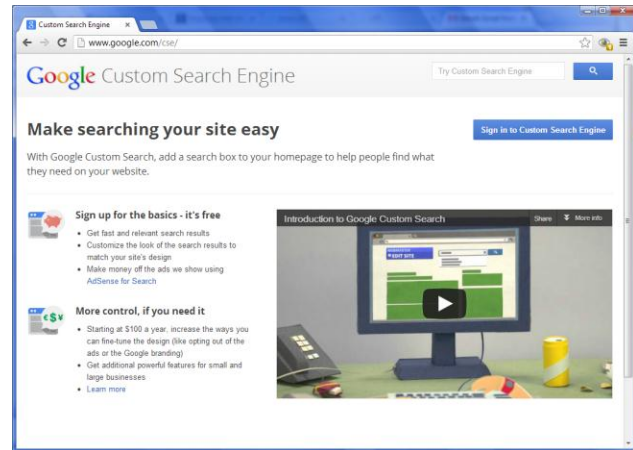


Fig. 1. Google custom search

The next step is making the profile of custom search engine that we're going to use in seaching and indexing the digital library systems. The making of custom search engine profile is quite easy by registering the URL (uniform resource locator) address of digital library into Google custom search engine pages, and after that Google will tell us how many pages that have been indexed on that address. The bigger amount of pages that have been indexed by Google the better digital library systems that we have. It tells that digital library systems is well structure, open access pages, bigger collections, stabil and reliable in the internet. This is the example of custom search engine profiles that have been made, as shown in Figure 2.

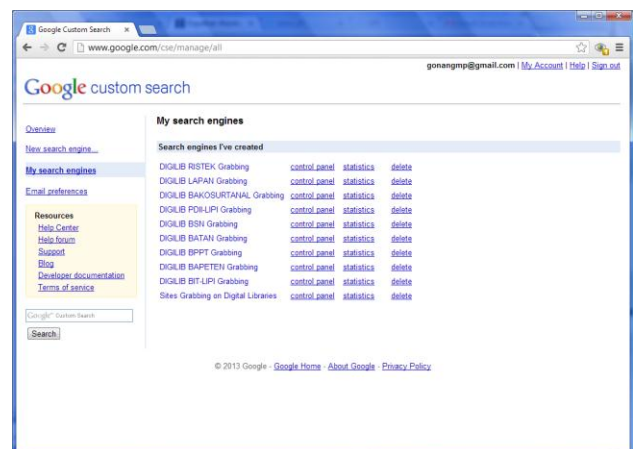


Fig. 2. Google custom search engine profiles

One thing that should be noted in Google custom search engine we can make so many profiles that index only one URL address or many URL addresses. If more than one of digital library systems that we want the results on the same knowledge mining tools, so we can use one profile of Google custom search for 2 URL addresses that represent each of digital library. After creating profiles for one or more digital library systems, Google custom search engine will give API code for every profiles that have been made. The API code which is provided by Google is to be used for calling Google custom search engine from digital library systems that have been attached by Google's API code. The example of API code is shown below:

```
<script>
(function() {
var cx = '013107692135569882287:-bj5nhhfrm';

var gcse = document.createElement('script'); gcse.type =
'text/javascript'; gcse.async = true;
gcse.src = (document.location.protocol == 'https:' ? 'https:' :
'http:') +
 '//www.google.com/cse/cse.js?cx=' + cx;

var s = document.getElementsByTagName('script')[0];
s.parentNode.insertBefore(gcse, s);

})();
</script>
<gcse:search></gcse:search>
```

Those script code is going to be used by digital library to do knowledge digging on specific URL address which represent digital library systems. How to attach that code into digital library systems, we need to enter the source code of digital library systems and deciding where to put Google's API code. By using the source code of digital library systems which was built on specific web programming language like PHP, ASP, Java, and others, and called the script which also included in digital library source code. Then from this, the systems will try to communicate with Google custom search engine servers to identify the custom search engine that was made and profile that attach to it. After Google custom search engine processing the request from digital library then sending back the results to digital library again to be showed in its window/panel. That way makes the knowledge finder tools seem to be part of digital library systems, because it shows the result on the same window/panel [13]. The structure of using Google custom search as knowledge finder in digital library is shown in Figure 3 below.

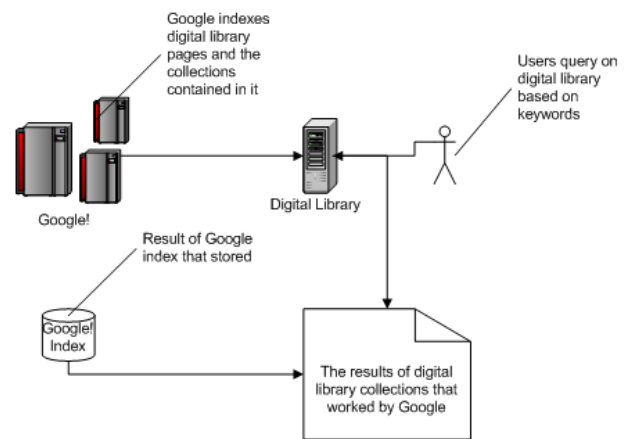


Fig. 3. Design of utilizing Google custom search

When users made query search on digital library based on keywords that inputted, just like when they used Google search engine. In count of seconds or less depends on digital library infrastructure and the speed of users internet, the better of that two aspects the better too the speed of results will show up. From the results seem like that knowledge finder tools is part of digital library systems, although as a structure that tools is part of the systems but as whole systems that tools is separated from the systems itself, because knowledge finder tools as systems is provided by Google custom search engine. No matter how the systems work, the sure thing is Google custom search engine that used as knowledge finder of digital library has showed tremendous things by showing the results of digital library collections with Google's style. Just like in Figure 3 showed that after users make query on digital library, the digital library then try to connect with Google based on the script that Google provided. After that Google send back the result to digital library to be showed on its display. When users clicked on one the results it will turn to one of data collections that digital library has. That's it, as simple as that the systems work and relation between Google custom search engine and digital library systems.

The next discussions will be how to make knowledge finder tools on digital library that utilize Google custom search. By attaching Google custom search engine code to any pages in digital library systems, if the process of attaching those code in right way, then the page will show up with keywords search box like we use Google search engine as usual, like was shown in Figure 4 below.

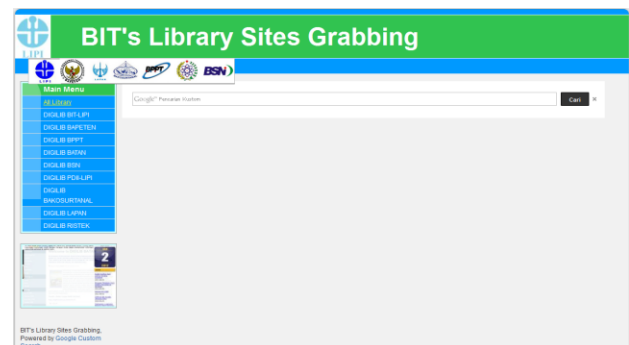


Fig. 4. Google custom search engine in digital library systems

In the center of digital library systems which was shown in Figure 4 is the example of Google custom search that fully utilized by digital library systems. The systems above was developed in order to search and show/grab the results based on some digital library systems which have been made the profiles on Google custom search sites. So, the digital library above will do search and show the result by users search keywords on digital libraries that run by institutions beneath Ministry Research and Technology (RISTEK) Republic of Indonesia. The institutions are Indonesian Institute of Sciences (LIPI), Nuclear Energy Regulatory Agency (BAPETEN), Agency for The Assessment and Application of Technology (BPPT), National Nuclear Energy Agency (BATAN), National Standardization Agency of Indonesia (BSN), Coordinating Agency for Surveys and Mapping Areas (BAKOSURTANAL), National Aeronautics and Space Institute (LAPAN), and Ministry of Research and Technology (RISTEK) itself. Every institutions that named above have central library and fortunately also have digital library systems which can be accessed via internet. Neither those digital libraries systems that run by those institutions embrace open access or semi-open access, but no one of those digital libraries in closed systems because Google will not recognize those systems. As described above in utilizing Google search engine as knowledge digger, the systems must be in open access or semi-open access to be recognized and grabbed the results.

The example of utilizing Google custom search above will search in all digital libraries systems that run by institutions beneath RISTEK, and also have capability in searching each digital libraries. The design of that systems in grabbing some digital libraries systems is shown below.

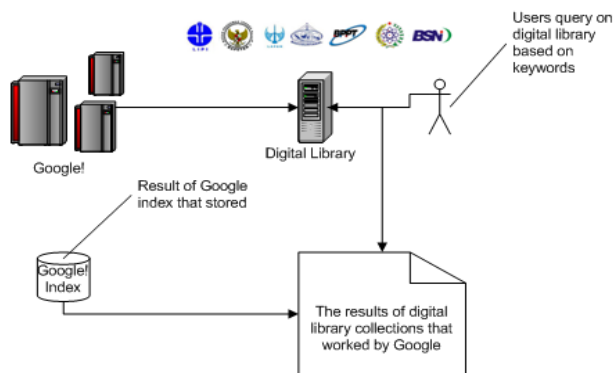


Fig. 5. Design of the systems in grabbing digital libraries

The systems above was designed as a combination between framework that was shown in Figure 3 and framework in Figure 5. The systems has ability to search and grab cross all over the digital libraries systems run by those institutions and also to search and get the results on each digital library [13]. The systems can work like this because in Google custom search engine sites the profile of all digital libraries systems and profile for each digital library have been created before like shown in Figure 2, so the systems just attach on each code that were given by Google. Each profiles that were created before will give the code to be attached to the systems. The systems above as shown in Figure 4, like it was just one application of digital libraries knowledge mining but actually

the systems have more than one applications because the systems have to attach more than one code that Google give for every profiles which have been created. So, the systems above have more than one applications that represent one profile for each applications. In All Library application even the profile is the combination of digital libraries under RISTEK that has more than one URL address, but they're included in one profile only and has one code only. Basically, if we open one of page that represent the institution beneath RISTEK, actually it's opening one application where inside of that application there is the code which provided by Google.

The example of how Google custom search working on a single application of digital library is shown below:

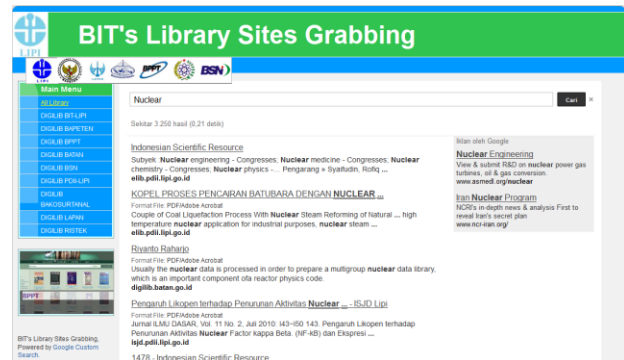


Fig. 6. The example of application in grabbing one digital library system

Each of digital libraries maybe have their own systems to search, index, grab, and show the result of queries that composed by users. Whether the systems are based on semantic web, indexing based on pages, or words query analysis, all the three methods have purpose in showing the best results of digital library collections for their users. The better techniques that used in knowledge mining digital library the better result that showed also based on user needs. Because the techniques that used by search engine have proven for its reliability, so the basic question is why don't we just utilize the technologies that search engine provided to be developed and customized based on each digital libraries needs. Who knows that using search engine technologies to index, search, and show the results will be better than we develop our own knowledge finder systems. Because Google search engine has the best techniques to index the webpages across the world, and we can use and customize it with free of charge.

Using Google custom search engine doesn't mean without any weakness, don't think that if we're using this technologies all the problems/solutions in knowledge finder of digital library will be perfectly fix. Like others systems that have strength and weakness factors, including Google custom search. The discussions above many told about the strength of Google custom search in helping us to build comprehensive tools to dig and search digital library systems. The next discussions will be focused on the weakness in using Google custom search engine [14], among them are:

- 1) Impossible to customize, because the nature of this technologies is Google's product, so developer could just make an API programs to call the function of this product. But it's limited in changing the function of

product. If we want to add searching options, like in the fields of Title, Author, Date Published, and so on... We just can't do that because Google index digital libraries and others websites with its own way. So Google doesn't index the website based on Title, authors, and so on, but Google index the whole/entire pages that Google found on single web. Of course this will be problems if we want to search in specific criteria with very large database. The best way to tackle that problems is by developing its own search and knowledge engine to search and find across the entire digital library collections. Of course this action has greater consequences because we have to invest in human resources, capital, and equipments.

- 2) Takes time to index the collections, if digital library has collections that just added it will be impossible to search that collections by using Google custom search in short time. Because Google needs time to index the collections and the same treat will be applied with others millions or even billions web pages across the world, and Google has schedule to index webpages. In this case, Google custom search that used as knowledge finder in digital library will be suit as knowledge finder only, it will not suit to be used as monitoring and controlling database collections.
- 3) Depends on the internet, as Google product of course this service is depend on the internet connection. It cannot operate in local network, even if Google custom search is used in the area where digital library is hosted it's still demand for internet connection. This condition occur because the topology of digital library that utilize Google custom search as knowledge finder is based on Figure 3, where the system will check to Google server after the request was done and will bring the results to the systems again by Google. This mechanism needs internet connection of course. This limitation has several impacts they are: needs longer time in processing and showing the results and if the digital library systems have failure in internet connection, user cannot use this knowledge finder tools which provided by Google [15].
- 4) Works in open systems or semi-open systems only, many digital libraries are adopting close systems which mean they protect collections only for those who have accessed to utilize that. If we're going to use Google custom search to be knowledge finder for digital libraries, it won't work for close systems. Simply because Google need to access and read the collections which lie on digital library. In close systems this will not going to happen because Google can't access that collections, and will give the result return to zero. If we're going to use Google as knowledge finder at least the digital library is adopting semi-open systems, which will give Google chance to recognize and index the collections. Even the collections are not in full open/access mode, but it will index metadata of the collections and this metadata still be accessible by users. If they want full collections, they have to become members of digital library that have grant access to that collections.

Results from analysis of using knowledge finder

technologies in digital library, one of that is provided by Google search engine through Google custom search and the other is by developed by its own. The summarized of comparative study is shown by following table.

TABLE I
COMPARATIVE STUDY ON KNOWLEDGE FINDER IN DIGITAL LIBRARY

| Strength / Weakness | Knowledge Finder Technologies | |
|---------------------|---|--|
| Strength | Build by own | Google custom search |
| | ✓ Easy to customize | ✓ Powerful search |
| | ✓ Unite with digital library applications | ✓ Easy to use like Google search |
| | ✓ Can be run locally (not depends on the internet) | ✓ No limit of digital libraries systems to be included |
| Weakness | ✓ Works in close & open systems | ✓ Can be integrated with others Google's products |
| | – Could be less powerful than search engine do in searching | – Impossible to customize |
| | – Vulnerable to be hacked | – Takes time to index |
| | – Depends on capabilities of human resources in development | – Un-independent |
| | | – Works in open or semi-open systems only |

From the Table I the dominant factor in strength of knowledge finder build by own is easy to customize, because the development of this kind knowledge finder connect directly to database collections of digital library. It will has ability to use that database to be searched what ever the developers want, mixing between fields, tables, or even databases. On other hand, the dominant weakness factor from this method is cannot provide the powerful results, because usually the development of this method was just mixing between fields, tables, and databases. So, it was lack with the pattern of searching, like the most searched articles, the most clicked articles, and even the most cited articles. This method was matching between the keywords that just inputted with the text inside the databases.

While Google custom search offers what knowledge finder that build by own can't do, Google custom search can do that with less effort from the developers. The strength and weakness from this method have described previously. So, when the purpose of knowledge finder in digital library is to provide satisfying results. Based on comparative study knowledge finder provided by Google custom search has more advantage than build by own, because it will provide with more satisfying results and less effort to make that happen.

IV. CONCLUSION

Using Google custom search as knowledge finder/digger tools for digital library is the most interesting and easy to create comprehensive and powerful tools to index and grab the collections that lie in digital library. By using Google custom search, web developers don't have to build knowledge finder

from very early phase of software development. All they have to do is building the page that going to be used as bridge to connect to Google server in showing the results of Google index based on user keywords input.

Utilizing Google custom search engine as knowledge finder tools in digital library will be different from knowledge mining that developed by own. The use of Google custom search in digital library just the same as using Google search engine, the tools only provide one box of search that can be inputted by users keywords to search the collections. The result that showed by the tools also looked like the results that showed by Google. The benefit from using Google custom search engine is the technologies that used by Google is advanced technology in searching data and information across all the web that can not catch by any others search engine, so the results which also provided by Google in knowledge finder pages are the best results. Even the results by using this tools can have better results compare to knowledge mining developed by own, because Google uses its crawling methods on digital library based on the trend, popularity, metadata, and others to find the best results for users.

Utilizing Google custom search engine as knowledge finder doesn't mean the solution in building perfect knowledge finder tools has been solve. Using this Google product as knowledge finder tools also has limitation beside the dependable on Google technologies which lead us to un-independent, also Google as a profitable company will insert business purpose in products which can be used freely by anybody. Here is the limitations by using Google custom search:

- 1) Impossible to customized, will be extremely impossible to modified Google custom search as knowledge finder except for which digital library that going to be used this tools.
- 2) Takes time to index the collections, the collections that have just been added to digital library cannot directly search by this tools.
- 3) Depends on the internet, because Google is products and company that very attached to internet, in providing its services also depends o the internet. If no internet connection impossible to utilize this tools.
- 4) Works in open systems or semi-open systems only, Google will index and available to search only if the digital library is not in close systems.

Based on comparative study knowledge finder provided by Google custom search has more advantage than build by own. No matter the choosen of knowledge finder tools in digital library that going to be used, one thing for sure it will help users to find and can be maximizing the results that hide on digital library collections. Between one tools and others tools, each has strength and weakness depends on the needs by digital library in fully utilizing digital library.

ACKNOWLEDGMENT

Author would like to thank to Head of Information Technology Division – Indonesian Institute of Sciences (LIPI) Mrs. Dianadewi Riswantini, for all the supports and

encouragements.

REFERENCES

- [1] Witten, I. H., Bainbridge, D., Nichols, D. M., "How to Build a Digital Library", Elsevier, USA, 2010.
- [2] Suleman, H., Fox, E. A., "A Framework for Building Open Digital Libraries", D-Lib Magazine, December 2001, Vol. 7 Number 12, 2001.
- [3] Arms, W. Y., "Digital Libraries", Massachusetts Institute of Technology, USA, 2001.
- [4] Marchionini, G., Plaisant, C., Komlodi, A., "Interfaces and tools for the library of congress national digital library program", Information Processing & Management, Vol. 34, Issue 5, September 1998, pp. 535–555, Elsevier, USA, 1998.
- [5] Rusbridge, C., "Towards the hybrid library", D-Lib Magazine, July/August 1998, UK, 1998.
- [6] Rossa, R., Sennyey, P., "The Library is Dead, Long Live the Library! The Practice of Academic Librarianship and the Digital Revolution", The Journal of Academic Librarianship, Vol. 34, Issue 2, March 2008, pp. 145–152, Elsevier, USA, 2008.
- [7] Nasukawa, T., Nagano, T., "Text analysis and knowledge mining system", IBM Systems Journal, Vol. 40, Issue 4, pp. 967 - 984, 2001.
- [8] Arms, W. Y., Bianchi, C., Overly, E. A., "An Architecture for Information in Digital Libraries", D-Lib Magazine, February 1997, Corporation for National Research Initiatives, USA, 1997.
- [9] Barroso, L. A., Dean, J., Hölzle, U., "Web search for a planet: The Google cluster architecture", IEEE, Vol. 23, Issue 2, pp. 22 - 28, March-April 2003.
- [10] Notess, G. R., "Custom Search Engines: Tools & Tips", Computers in Libraries, v28 n5 p16-17 May 2008, Information Today, Inc., USA, 2008.
- [11] Vasumathi, D., "Web mining using pattern discovery techniques", Indian ETD Repository @ INFLIBNET, Jawaharlal Nehru Technological University, URI: <http://hdl.handle.net/10603/2194>, Issue Date: 8-Aug-2011.
- [12] Thelwall, M., Binns, R., Harries, G., Page-Kennedy, T., Price, L., Wilkinson, D., "Custom interfaces for advanced queries in search engines", Aslib Proceedings, Vol. 53 Issue 10, pp.413 - 422, 2001.
- [13] Perdananugraha, G. M., "KNOWLEDGE MINING IN DIGITAL LIBRARY", International Conference on Information Technology and Electrical Engineering (CITEE), Yogyakarta, Indonesia, 2012.
- [14] Bradley, P., "Human-powered Search Engines: An Overview and Roundup", Web Magazine for Information Professionals, Issue 54, link: <http://www.ariadne.ac.uk/issue54/search-engines>, 2008.
- [15] Ryan, P. A., "Case 12: Google Inc. The Future of the Internet Search Engine", Strategic Management and Business Policy, link: <http://aykancandemir.org/str/google.pdf>, 2010.



Gonang May Perdananugraha, ST., MT. was born in Temanggung, Central Java Province, Indonesia on May 2, 1982. Gonang is currently work as researcher at Indonesian Institute of Sciences (LIPI), in the field of information technology with specialization in e-government, web-based technologies, information systems, web-based open source software, web-based security, digital library, knowledge mining, and socio-cultural information technology. Gonang has graduated from Indonesia Institute of Technology (ITI) in major of informatics engineering and Pelita Harapan

University (UPH) in major of Master of Industrial engineering. As young researcher Gonang has several scientific articles that published in national and international journals and proceedings. Gonang is also candidate for Ph.D student at one of the top University in Australia.

