
MANAGEMENT APPROACH TO ONLINE ENCYCLOPEDIAS BY USING GOOGLE CUSTOM SEARCH ENGINE VIA GOOGLE CO-OP

BADAN BARMAN

Abstract

List and discusses the online general purpose encyclopedia that are available on public domain. All these encyclopedias are used in designing a Google Custom Search Engine via Google Co-Op, so that user can search in a single point and all relevant results can be displayed on that platform by searching all the encyclopedias individually. By this way it hopes to manage the entire online encyclopedias that are available on public domain. This custom search engine can be linked with other institutional website / blog / wiki etc for answering some basic reference question or finding background information on any topic (?).

Keywords: Knowledge management/ Open access encyclopedia/ Online encyclopedia/ Reference desk/ Google Custom Search Engine

1. Introduction

Encyclopaedia is a circle of knowledge, a work that represents synthesis of knowledge or compendium of knowledge. It attempts to bring some order to the knowledge reflecting state of knowledge, as it exists during the period of its compilation. The included knowledge is related to kind of readership that an encyclopedia intends to serve. The general Encyclopedias often contain guides on how to do a variety of things, as well as embedded dictionaries and gazetteers.

Previously the encyclopedias were published in printed form and it has several drawbacks such as written by single writers, to make space for modern topics, valuable material of historic use regularly had to be discarded etc. The need to overcome such barriers and the new generation need, increase use of IT, advent of internet and World Wide Web all favour for publication of encyclopedia on some more sophisticated medium, firstly in the form of CD-ROMs for use with personal computers and later in the form of online encyclopedia. Further the interactive nature of the Internet allowed the collaborative generation of encyclopedias by the expert of the subject and providing free access to the public from all over the world. Such collaborative project includes Nupedia, Everything, Open Site, Wikipedia etc, some of which allowed anyone to add or improve content.

While the printed encyclopedia follows the alphabetical arrangement or organization by hierarchical categories, the fluidity of electronic media and internet allows multiple methods of organization of the same content with previously unimaginable capabilities for search, indexing and cross reference. They are now become larger compendia, includes all that came before them, articles were supplemented with video and audio files as well as numerous high-quality images, animations etc., which were impossible to store in the printed format. Hyper linking between conceptually related items is also a significant benefit. On-line encyclopedias, like Wikipedia, offer the additional advantage of being (potentially) dynamic: new information can be presented almost immediately, rather than waiting for the next release of a static format (as with a disk-or paper-based publication), it also allow searching through article text for any keyword(s).

2. The need

When user seek information over internet by searching in the top most popular search engines Google, Yahoo, MSN or like other, they frequently loss themselves in the vast amount of irrelevant information. The search engine also at times does not give more emphasis on the online encyclopedia, those that are available on public domain in citing background sources of information. At the same time to find article on a specific online encyclopedia one might search each encyclopedias individually, a way of losing valuable time. All such problem urges the need of another search engine that primarily focused in retrieving background information on any topic by searching each of the encyclopedias individually.

3. Methodology

To compile a list of general purpose, open access encyclopedias I conducted a number of web searches in the three 'top most popular search engines' i.e. Google, Yahoo and MSN by using different key words. The search engines which I have mention as "most popular" is not my personal views, but the results of different online news (such as <http://news.bbc.co.uk/1/hi/magazine>; <http://searchenginewatch.com/>; <http://www.manchestersonline.co.uk/men/news>; <http://seroundtable.com/archives> etc.). I also appreciate that simply using a top class search engine to look for a list of online encyclopedias will not necessarily lead to quality and relevant results so future provision is also made to incorporate the online encyclopedias to the list of site to be search by the custom search engine as it comes to the focus. Out of a big list, selection is made to identify the open access general purpose encyclopedias, in selection three points were consider –Open access or closed access, Scope and coverage. The selected encyclopedias are then use to build a custom search engine via Google Co-Op.

4. Findings

The online general purpose encyclopedias that are find their way to the public domain can be categorize as follows

4.1 Digitization of old content

The online encyclopedia firstly finds their way to the public domain by digitization of old printed encyclopedias. Such encyclopedias includes the following-

Project Gutenberg (www.gutenberg.org/etext/200): Project Gutenberg (often abbreviated as PG) is a volunteer effort to digitize, archive, and distribute cultural works. Michael Hart founded it in 1971; it is the oldest digital library. Most of the items in its collection are the full texts of public domain books. The project tries to make these as free as possible, in long-lasting, open formats that can be used on almost any computer. As of October 2006, Project Gutenberg claimed over 19,000 items in its collection, with an average of over fifty new e-books being added each week. In January 1995, Project Gutenberg started to publish the ASCII text of the Encyclopaedia Britannica, 11th edition (1911), but disagreement about the methods halted the work after the first volume. For trademark reasons this has been published as the Gutenberg Encyclopedia. Project Gutenberg has restarted work on digitizing and proofreading this encyclopedia but no records were found whether it is published or not.

LoveToKnow (http://www.1911encyclopedia.org/Main_Page): The LoveToKnow is a free Online Encyclopedia that based on what many consider to be the best encyclopedia ever written: the eleventh edition of the Encyclopaedia Britannica, first published in 1911. The Eleventh Edition filled 29 volumes and contains over 44 million words. It contains over 40,000 articles written by over 1,500 authors within their various fields of expertise. LoveToKnow is giving all the entries of it, preserving the treasured entries that make it so unique, and where necessary and possible adding the current point of view. On each subject there are in fact two pages. One is an article from the 1911 Encyclopaedia Britannica or its 1922 additional material that is to be found with the “article” tab at the top of the page. The other is for more modern additions, corrections and comments made by readers which can be found with the “what’s new” tab. In 2002, ASCII text of all 28 volumes of Encyclopaedia Britannica was published on <http://www.1911encyclopedia.org/> domain.

Encarta (<http://encarta.msn.com/>): Encarta is the first digital multimedia encyclopedia published and updated frequently by Microsoft Corporation. Microsoft initiated Encarta by purchasing non-exclusive rights to the Funk and Wagnalls Encyclopedia, incorporating it into its first edition in 1993. In the late 1990s, Microsoft bought Collier’s Encyclopedia and New Merit Scholar’s Encyclopedia from Macmillan and incorporated them into Encarta. Thus the current Microsoft Encarta can be considered the successor of the Funk and Wagnalls, Collier, and New Merit Scholar encyclopedias. None of these formerly successful encyclopedias are still in print. As of 2005, the complete English version, Encarta Premium, consists of more than 68,000 articles, numerous images and movies, and homework tools, and is available on the World Wide Web by yearly subscription or by purchase on multiple CD-ROMs or DVD-ROM. For years, Encarta came in three primary software flavors: Basic, Premium, and Reference Library (price and features in that order). Many articles can also be viewed online for free, a service supported by advertisements. Microsoft’s Encarta was a landmark example, as it had no print version.

HighBeam™ Encyclopedia (<http://www.encyclopedia.com/>): Probably the most important and successful digitization of an encyclopedia was the Bartleby Project's online adaptation of the Columbia Encyclopedia, Sixth Edition, 2000, which was put online at <http://www.bartleby.com/65/> in early 2000. It is the Internet's premiere free encyclopedia. It provides users with more than 57,000 frequently updated articles from the Columbia Encyclopedia, Sixth Edition. Each article is enhanced with links to newspaper and magazine articles as well as pictures and maps which all provided by HighBeam Research.

Easton's Bible Dictionary, 1897 (<http://eastonsbibledictionary.com/>): Easton's Bible Dictionary generally refers to the Illustrated Bible Dictionary, Third Edition, by Matthew George Easton M.A., D.D. (1823-1894), published three years after Easton's death in 1897 by Thomas Nelson. Because of its age, it is now a public domain resource. The Christian Classics Ethereal Library digitized it. Easton's Bible Dictionary contains an extensive set of entries used in the Bible, from a 19th century Christian viewpoint. Some of the entries in it are now out-of-date; many are only short dictionary entries. Despite its name, many of the entries in Easton's are encyclopedic in nature, though there are short, dictionary-like entries. It contains nearly 4,000 entries.

4.2 Creation of new content

In recent years another related branch of activity to be considered in building online open access encyclopedias are the creation of new, free contents on a volunteer basis and it itself can be divided into following two broad categories

4.2.1 Encyclopedias those are no longer online :

- a. Interpedia: The idea to build a free encyclopedia using the Internet can be traced at least to the 1993 Interpedia proposal; it was planned as an encyclopedia on the Internet to which everyone could contribute materials. Rick Gates, who reportedly posted it on 22 October 1993 to alt.internet.services Usenet newsgroup, initiated it. R. L. Samuell, a participant in early discussions on the topic, coined the term Interpedia. Discussion initially took place on a high-volume mailing list, and in November 1993 the Usenet newsgroup comp.infosystems.interpedia was created. But the project never left the planning stage and finally died, taken over by the explosion of the World Wide Web, the emergence of high-quality search engines, and the conversion of existing material.
- b) Nupedia (<http://nupedia.8media.org/>): Nupedia was Web-based free peer reviewed encyclopedias whose articles were written by experts and licensed as free content. It was founded by Jimmy Wales and underwritten by Bomis, with Larry Sanger as editor-in-chief. Nupedia lasted from March of 2000 until September 26, 2003. Wikipedia, the predecessor of Nupedia, has since assimilated much of its content. An extensive peer-review process designed to make its articles of a quality comparable to that of professional encyclopedias characterized Nupedia. Nupedia wanted scholars to volunteer content for free. Before it ceased operating, Nupedia produced 24 articles that completed its review process (three articles also existed in two versions of different lengths), and 74 more articles were in progress.

- c) GNUpedia: An initiative, which did not come to fruition but had some interesting philosophy.

4.2.2 Currently online encyclopedias

- a) Project Galactic Guide (<http://www.galactic-guide.com/>): Project Galactic Guide is one of the oldest Internet encyclopedia projects. Project Galactic Guide (PGG) is a collaborative project inspired by The Hitchhiker's Guide to the Galaxy, the fictional encyclopedia in books by Douglas Adams. Participants of the alt.fan.douglas-adams newsgroup started the project in 1991. Paul Jason Clegg proposed the idea of creating a real guide to the galaxy. It is a collaborative Internet attempt at creating an electronic reference meant to guide and misguide its readers in matters of life, death, and finding a parking space anywhere in the universe. PGG has three classifications for articles: real, unreal and semi-real. Real articles attempt to provide accurate, factual information in an amusing and enjoyable style. Unreal articles are for humor value only. Semi-real articles are a mix of both. All PGG articles have their classifications clearly marked their titles to avoid confusing the reader. The project's website is currently maintained by Alex McLintock. There are 1780 articles in the Project Galactic Guide but no new articles have been added since 2000 and Work on the guide has slowed to a halt in recent years and hope to died very soon, this is probably partly due to the development of h2g2, a more official project along similar lines, which was backed by Douglas Adams and his company, The Digital Village.
- b) Everything2 (<http://everything2.com/>): The predecessor of E2 was a similar database called Everything (later labeled "Everything1" or "E1") which was started around March 1998 by Nate Oostendorp and was initially closely aligned with and promoted by Slashdot. The E2 software offered vastly more features, and the Everything1 data was twice incorporated into E2: once on November 13, 1999 and again in January of 2000. Everything2, or E2 for short, is a collaborative Web-based community consisting of a database of interlinked user-submitted written material. E2 is moderated for quality, but has no formal policy on subject matter. Writing on E2 covers a range of topics one cannot imagine, but there is also humor, poetry, fiction, opinion, criticism, personal experiences and other things that are simply hard to categorize. Only logged-in E2 users create entries and add information in E2 and only the author of an article or an editor appointed by the site administrators can edit an article. E2 categorizes writeups (=article) into four types: person, place, idea, and thing. Articles are written in a simplified HTML dialect and do not contain images. The copyright in an article rests with the author, and writing enters into no agreement to any kind of license on E2 (except for giving the site permission to publish). Authors retain the right to place their work in the public domain, to release it under a copyleft license such as one of those offered by the GNU project or Creative Commons, or to request the removal of their work from the site at some later date. Now the University of Michigan, located in Ann Arbor, has hosted Everything2 and it is a privately owned Company (Blockstackers Intergalactic Company) but does not make a profit.
-

-
- c) Wikipedia (<http://en.wikipedia.org>): Wikipedia was launched as an English language “.com” project on January 15, 2001 as a compliment to the expert written and now defunct Nupedia. In 2002 “.com” was deprecate in favor of “.org”. Now the non-profit Wikimedia Foundation operates it. Larry Sanger and Jimmy Wales created it, Wales has described Wikipedia as “an effort to create and distribute a multilingual free encyclopedia of the highest possible quality to every single person on the planet in their own language. Wikipedia produced enormous amounts of content, thousands of excellent articles, and is getting high profile, international recognition as a new way of obtaining at least a rough and ready idea about very many topics. Its surprising success may be attributed, briefly, to its free, open, and collaborative nature. It currently contains 1,468,966 articles; content licensed under the copyleft GNU Free Documentation License. However Wikipedia’s articles may be considered to be of a trivial nature. Wikipedia main servers are in Tampa, Florida with additional server in Amsterdam and Seoul.
- d) Open Site (<http://open-site.org/>): The Open-Site project is a free, trusted internet encyclopedia founded in 2002 by Michael J. Flickinger in an effort to build a free categorized community-built encyclopedia, inspired by the Open Directory Project. Open-Site has the characteristic of both a dictionary and encyclopedia. The best suitable description for Open-site is “informational, volunteer-based web-site”. It offers both definitions on words and encyclopedic articles on topics, yet also include a huge range of other information. The Open Site Encyclopedia Project aims to be the most comprehensive, informational, volunteer-powered web site on the Internet. Its purpose is to create, into an organized structure, content of different types, from different topics, and to share it back to the Internet community. Open-Site is edited by volunteer editors and accepts content submissions from the public to provide accurate information. The project currently has no budget with Android Technologies, Inc. providing a dedicated server in the state of Texas in the United States.
- e) h2g2 (<http://www.bbc.co.uk/dna/h2g2/>): The title The Hitchhiker’s Guide to the Galaxy, the best-selling book by Douglas Adams is often abbreviated as “HHG”, “HHGG”, “HHGTTG”, or “H2G2”. The series is also often referred to as “The Hitchhiker’s Guide”, “Hitchhiker’s”, or simply “[The] Guide.” h2g2, was launched in April 1999 as the “Earth Edition” of The Guide by the author of the series, Douglas Adams, and his friends and colleagues at The Digital Village, and the BBC took over the running of the site in February 2001 as part of their drive to develop new and innovative online services. When the site became part of BBCi, the BBC insisted on moderating all contributions to the site soon after they were made. h2g2 is really two separate but complementary Guides, one Edited and one Unedited. The Edited Guide forms only a small part of h2g2 as a whole and consists of articles (usually called ‘Entries’) which have passed through a peer review process, and then been checked and tidied up first by a volunteer sub-editor and then, more briefly, by an in-house editor. On August 16, 2006, h2g2 added its 8,000th entry to the edited guide. Most of the site’s ‘cultural life’ takes place in the far larger Unedited Guide, which contains, amongst other things, various clubs and societies, discussion areas, Researchers’ h2g2 homepages (known as
-

their ‘personal spaces’), and writing workshops. It is a website now featuring a range of an unconventional guide to life, the universe and everything, an encyclopedic project where entries are written by people from all over the world. It has thousands of entries on all sorts of subjects. The result is a living, breathing guide that’s constantly being updated and revised, driven forward by the very people who use it. The project work on the same line as that of ‘Project Galactic Guide’.

- f) The Citizendium Project (http://pilot.citizendium.org/wiki/Main_Page): The Citizendium (sit-ih-ZEN-dee-um), a “citizens’ compendium of everything,” is a non-profit, open content wiki project that combines public participation with gentle expert guidance. Larry Sanger, is the founder of this project and his title is Editor-in-Chief. The project begins its life with all of Wikipedia’s articles, and then people start making changes to articles in the Citizendium. On a very regular basis, it refreshes the copies of Wikipedia articles. If an entry in the Citizendium remains unchanged since being copied from Wikipedia, but the Wikipedia version has, then it uploads the most recent Wikipedia article. But if the Citizendium has changed an article, then it is not refreshed. In this project personal accountability including the use of real names for logging and more is expected to create a responsible community and a good global citizen. The Citizendium Foundation, Mount Hermon, runs it.

5. A management approach to online general purpose, open access encyclopedias

5.1 About Google Custom Search Engine Via Google Co-Op

Google has launched their flavor of a Rollyo, or Eurekster or Yahoo Search Builder by named Google Custom Search Engine. It is built into the Google Co-Op, which is a platform that enables one and to offer other a customized version of its search engine, narrowing down its vast index so that the results are more relevant to users. The platform can be used to direct the search engine to search some particular sites such as ones own blog / website / wiki etc. as such also can be used to build a free, highly specialized ‘Customize Search Engine’ for a particular need or approach that will reflects ones knowledge and interests. In reverse by using this facility one can also exclude sites that he/she does not want in the results: a way that can be used to throw up spam sites to be excluded from search results. The Google Co-Op can be use to

1. Specify or prioritize the sites one wants to include in searches.
 2. Build a custom search engine.
 3. Invite other community to contribute to the search engine.
 4. Place the custom search engine and search results on a particular website / blog / wiki etc.
-

5.2 Creating a new search engine

The Google Platform 'Google Co-Op' is used in building a Custom Search Engine by name 'Reference Desk'. The basic background information regarding the reference desk is.

1. Search engine name: 'Reference desk'.
2. Search engine description: 'It will search all general purpose internet encyclopedias'.
3. Search engine keywords: 'Internet encyclopedia, Reference desk'
4. Search engine language: 'English'
5. Sites to search: All the above mention encyclopedias are included in the list of sites to be search by 'Reference Desk'. Some of them are
 "http://en.wikipedia.org
 http://encarta.msn.com/
 http://www.bartleby.com/65/
 http://www.encyclopedia.com/
 www.gutenberg.org/etext/200'
6. How to search these sites: The search engine 'Reference Desk' will "Search only these sites" (The list that mention above).
7. Collaborate with Others: Any one may volunteer to add more sites to this custom search engine. However to closed the door to web spam, which will soon come in, choosing the option 'Only people I invite may contribute to this search engine' imposes the restriction. The provision by virtue of its nature is there to add more sites by going to the search engine's control panel or while surfing the web by using Google Marker.
8. The code for linking: Google provides the search engine with a homepage. In addition, one can also place a search box and /or search results on his / her own website in one of the several ways that it provides.
9. Address of the 'Reference Desk' search engine: Below the URL of the Google Custom Search Engine 'Reference Desk' is given. <http://www.google.com/coop/cse?cx=005330743865816127897%3Aaxzib8gpd6j4>

6. Conclusion

Day by day more and more encyclopedias are going to the public domain. Building a customize search engine hope to be the cleanest technology to manage the entire online open access, general-purpose encyclopedias. It will also save the time from the user, help in retrieving a list of high relevant encyclopedic article, act just like's one stop shopping mall in retrieving background information and so on. The code can be used to further edit this search engine, and placed the search box to one's own

educational website / blog / wiki etc for finding background information on any topic (?). The student of any discipline can use it as a reference desk to identify the relevant resources, to complete their class assignment or for self-study. All library and information science professional can use it in providing reference service to the user almost instantly. Libraries can make use of this search engine by making it readily available to their patrons through the libraries home page by way of linking.

References

1. Bartleby Project's (<http://www.bartleby.com/65/>)
2. Citizendium (<http://www.citizendium.org/>)
3. Citizendium Project (http://pilot.citizendium.org/wiki/Main_Page)
4. Easton's Bible Dictionary, 1897 (<http://eastonsbibledictionary.com>)
5. Encarta (http://encarta.msn.com/artcenter_/browse.html)
5. Encarta® (<http://encarta.msn.com/>)
6. Everything2 (<http://everything2.com/>)
7. Google Co-Op (<http://www.google.com/coop/cse/>)
8. h2g2 (<http://www.bbc.co.uk/dna/h2g2/>)
9. HighBeam™ Encyclopedia (<http://www.encyclopedia.com/>)
10. hitchhikers (<http://www.bbc.co.uk/cult/hitchhikers/>)
11. LoveToKnow (http://www.1911encyclopedia.org/Main_Page)
12. Nupedia (<http://nupedia.8media.org/>)
13. Open Site (<http://open-site.org/>)
14. Project Galactic guide (<http://www.galactic-guide.com/>)
15. Project Gutenberg (http://www.gutenberg.org/wiki/Main_Page)
16. Project Gutenberg (www.gutenberg.org/etext/200)
17. Wikipedia (<http://en.wikipedia.org>)

BIOGRAPHY OF AUTHOR



Badan Barman holds MLISc degree from Gauhati University with 1st position, qualified UGC (NET) and presently working as a Research Assistant in the ICSSR Project “Rural Libraries of North East India: A Machine readable database” at DLISc, Gauhati University. He has attend a number of National Seminar and presented a paper in PLANNER 2006.

Email : badanbarman@gmail.com
