Genomics & Informatics Vol. 4(3) 125-128, September 2006

BioCC: An Openfree Hypertext Bio Community Cluster for Biology

Sungsam Gong[1], TaeHyung Kim[1], Jungsu Oh[1], Jekeun Kwon[1], SuAn Cho[1], Dan Bolser[2] and Jong Bhak[1]*

[1]Korean Bioinformation Center (KOBIC), KRIBB, Daejeon 305-806, Korea, [2]MRC-DUNN, Cambridge CB2-2XY, England, United Kindom

Abstracts

We present an openfree hypertext (also known as wiki) web cluster called BioCC. BioCC is a novel wiki farm that lets researchers create hundreds of biological web sites. The web sites form an organic information network. The contents of all the sites on the BioCC wiki farm are modifiable by anonymous as well as registered users. This enables biologists with diverse backgrounds to form their own Internet bio-communities. Each community can have custom-made layouts for information, discussion,and knowledge exchange. BioCC aims to form an ever-expanding network of openfree biological knowledge databases used and maintained by biological experts, students, and general users. The philosophy behind BioCC is that the formation of biological knowledge is best achieved by open-minded individuals freely exchanging information. In the near future, the amount of genomic information will have flooded society. BioCC can be an effective and quickly updated knowledge database system. BioCC uses an opensource wiki system called Mediawiki. However, for easier editing, a modified version of Mediawiki, called Biowiki, has been applied. Unlike Mediawiki, Biowiki uses a WYSIWYG (What You See Is What You Get) text editor. BioCC is under a share-alike license called BioLicense (*http://biolicense.org*). The BioCC top level site is found at *http://bio.cc/*

Keywords: openfree hypertext, bio-domains, omics, bioinformatics, biolicense

Introduction

The internet as we know it was formed in the early 1990's when Tim Berners-Lee distributed a daemon program called HTTPD (http://*www.w3.org/People/Berners-Lee/Longer.html*). HTTPD is a hypertext text processing daemon or server that runs on a hardware computer network (Berners-Lee *et al.*, 1994). In the mid 1990s, many web sites using the HTTPD and HTML (Hypertext Markup Language) format sprang up (*http://en.wikipedia.org/wiki/History_of_the_World_Wide_Web*). These sites were mostly static in that only the web masters can edit and add to the contents of the information provided by the web servers. However, as technology developed and the general public started to access the Internet, there was a higher demand for more up-to-date information and more knowledge exchange (Cunningham *et al.*, 2001; Stephen *et al.*, 2006). This resulted in a new system that can be called an "openfree hypertext web" (*http://en.wikipedia.org/wiki/Web_2*) (Kevin *et al.*, 2006).

The major difference between the open free hypertext and the existing web service is that the open free hypertext allows clients to add, delete, and edit web contents with minimum restriction. Often the web contents were also free (i.e., CopyLeft). This seems radically progressive and risky in information management. However, that is in fact closer to the early practice of using the Internet in the early 1990s. While openfree technology such as wiki was not widely available, the Internet became more restricted in the mid 1990s. One of the major fields of science that benefited most by the advance of the Internet has been biology (Guest *et al.*, 2003). Biological data are complex, diverse, and messy to handle. Also, human annotation is often a critical component of the biological data and its updates. Therefore, many large biological institutes have been running web sites with a remarkably open and free license scheme such as GNL (*http://www.gnu.org*). The majority of such data and databases has been free (Holger *et al.*, 2005; Kai *et al.*, 2006)In that open culture, there have been various community projects such as Bioperl, Biojava, Biopython, and Biolinux (*http://bioperl.net, http://bioperl.org, http://biojava.net, http://biojava.org, http://biolinux.net, and http://bioinformatics.org*). These individual projects have been linked by groups of researchers who advocate an open, free, and fast exchange of biological information. As one such openfree project, we present BioCC. BioCC is a top level portal site for open hypertext web sites that use Wiki (*http://wiki.org*). The concept behind BioCC is the ongoing construction of a large scale network of wiki-based

*Corresponding author: E-mail j@bio.cc,
Tel +82-42-869-4318, Fax +82-42-869-4310
Accepted by 4 August 2006

web sites. BioCC can be found at the following URL; *http://bio.cc/*. The history of BioCC and its diverse activities date as far back as 1996 with the first community project proposals for bioperl and biojava. BioCC's scope of activity is similar to those of *http://bioinformatics.org*, that has been successfully implemented some years later. BioCC, however, is different in its philosophy in that it is intended to form a networked cluster of very specialized wiki sites that have a common license, templates, and a philosophy for sharing, instead of becoming one single top level portal site such as *http://bioinformatics.org, http://google.com* and *http://yahoo. com.* BioCC maintains over 1000 biologically relevant internet domains that function as dynamically changing nodes for the whole cluster.

Overall, it forms a gigantic knowledge network database with many volunteers from diverse backgrounds. BioCC's development or evolution roadmap includes 1) a network of special knowledge domains, 2) a deep search engine that finds database entries as well as web pages, 3) an automatic word linking for an infinite number of word connections, 4) anautomatic renewal and weighting system for web site interconnection, and 5) an artificial intelligence knowledge query system for easier and contextual knowledge retrieval. below the BioCC level, there are high level portal sites that are more abstract than specialized domains such as *http://zincfinger. org.*

Methods

BioCC uses Mediawiki as the basis for developing its own WYSIWYG (What You See Is What You Get) wiki program called Biowiki. Mediawiki is based on the PHP (*http://php. net*) programming language which is flexible, manageable, and scaleable. Mediawiki has its own simple wiki grammar called wiki markup. The wiki markup is a set of syntax used to format and edit Mediawiki pages. Many users who do not know the wiki markup find it difficult to write a wiki web page. Although Mediawiki has its own editing tool, it provides a limited set of functions to format a variety of HTML codes. To solve this problem, the Biowiki wiki program integrates a graphical editor called FCKeditor (*http://www.fckeditor.net*) as its major editing tool. FCKeditor enables an easy and intuitive editing in a visual and straight forward WYSIWYG environment.

BioCC farm has thousands of domain names that point to a single or multiple server machines. It takes advantage of Apache (*http://www.apache.org*), one of the most common HTTP daemons, as a main server engine. To successfully handle hundreds of active Biowiki sites within a small number of machines, the "NameVirtualServer" Apache module has been adopted. BioCC has 2 dual core AMD Opteron 275 processors and 10GB RAM based on theFedoreaCore version 5 operating system with a 2.6.17 Linux kernel.

The virtual web sites (called bio-domains),such as bioperl.net, biocourse.org, biocorea.org, and biopeople. org, are diverted to the Apache web server's virtual server, and users outside access the bio-domains as distinct individual and international internet domains.

Results

BioCC, which is the mother of all the Biowiki operated sites, hosts hundreds of openfree hypertext sites for biology. Among them, we introduce the five most active Biowiki sites (Fig. 1). They are 1) BioCourse.org, 2) BioPedia.org, 3) BioSpecies.org, 4) BioPeople.org, and 5) BioCorea.org. Each of these sites is discussed in moredetail below. However, there are many other useful sites, for example, omics.org and Variome.org. Omics is the top level directory site for all the new-omicsdisciplines in biology such as genomics, proteomics, and interactomics. Variome.net is for SNP (Single Nucleotide Polymorphism) related research portals.

BioCourse.org (http://biocourse.org)

BioCourse is an open information archive designed for novice or intermediate level students and researchers in the field of bioinformatics, including general biology. There have not been many useful educational portal sites for bioinformatics which deal with a wide variety of subjects such as statistics, computer systems, genomics, microbiology, and programming languages. Internet users usually visit web sites of interest and manage them by using Bookmark utilities. However, this is tedious to maintain and update. Also, clients (internet users) cannot actively add, edit or remove the contents of such web pages. In order to overcome these limitations, we have developed BioCourse.org, an openfree web system for the exchange of learning and teaching materials. The purpose of BioCourse is to help students quickly grasp large amounts of information about research methods in various fields. The contents of BioCourse can be classified into six main categories:1) BioLanguage, 2) BioTool, 3) BioSystem, 4) BioDatabase, 5) BioLecture, and 6) BioJournal. BioLanguage deals with programming languages such as Perl, Java and Python. BioTool introduces biological and bioinformatics utilities, such as BLAST, and users can post their home-made utilities. BioSystem and BioDatabase deal with computer operating systems and Database Management Systems (DBMS). BioLecture contains introductory course materials

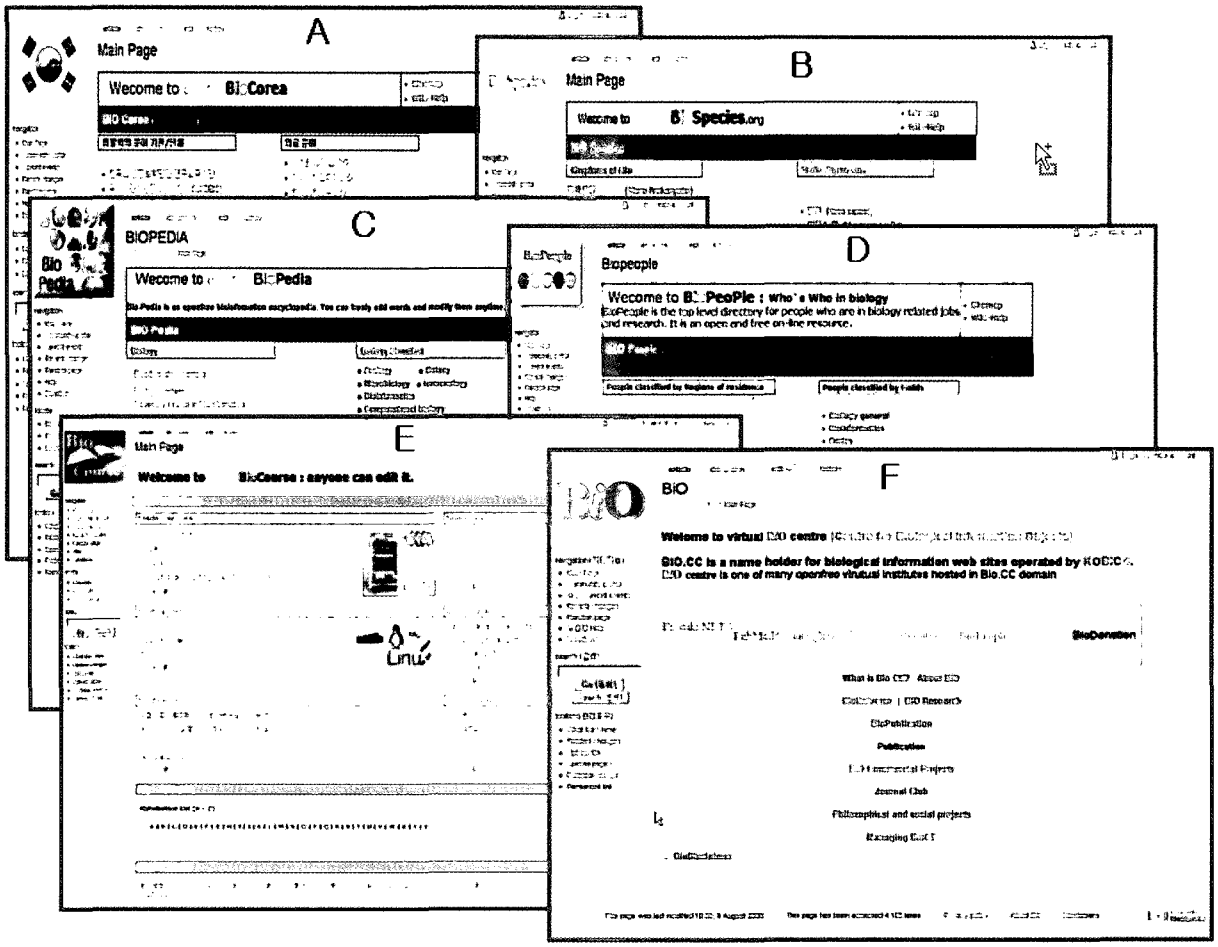

Fig. 1. Screen shots of BioCC and 5 Biowiki sites. A: BioCorea.org, B: BioSpecies.org, C: BioPedia.org, D: BioPeople.org, E: BioCourse.org, F: Bio.CC.

and BioJournal cites scientific journals in the fields of biology and bioinformatics.

BioPedia.org (http://biopedia.org)

BioPedia is an openfree encyclopedia dedicated to all biological glossaries and vocabulary. An enormous amount of terms and jargons in the filed of life science has been accumulated due to the rapid development of life sciences in the last couple of decades. Many novel terms, such as interactome and interactomics, have been coined in this -omics era. Even bioinformatics, genomics, and proteomics are fairly recent terms. However, such -omics jargon can create problems in communication among biologists due to ever-changing definitions. Hence, we developed an openfree web-based encyclopedia with the aim of keeping technical terms accurate, and can be modifiable in order to keep up with the rapid development in the life science. Biopedia will be utilized as a resource database for semantic web and ontology networks in biology.

BioSpecies.org (http://biospecies.org)

BioSpeices is an openfree directory service for listing all the species in the world. Its top level category has three major super kingdoms; prokaryote, eukaryote, and virus. It has amodel organisms section such as human, mouse, rat, C.elegance, and E.coli. BioSpecies is designed to satisfy the professional needs of biologists rather than general users. Most of BioSpecies's pages contain classification information from kingdoms down to subspecies, and a basic description of the organismsuch as scientific or common name, habitat, diet, genome size, and industrial productivity. Any newly identified organism which acquires an official authentication can be freely posted.

BioPeople.org (http://biopeople.org)

BioPeople is a web based who's who service in a biology domain. It aims to maintain information about scientists in the biology fields, providing a person's profile such as affiliation (s), research interests, collaborations, and publication information. It aims to be a voluntary community site. We have classified groups of scientists by region, research field, and affiliation, so that users are able to search for researchers' names by country, research field, or affiliation. We selected the top level directory for people who are in life sciences related jobs. Biopeople is a publicly open utility providing cyber-communication which gathers information worldwide for life scientists to help connect them to each other.

BioCorea.org (http://biocorea.org)

Biology is an extensive branch of science, consisting of many research centers, companies, and laboratories. BioCorea.org is an openfree portal site for scientists in Korean life science fields. BioCorea.org has been developed to facilitate communications among local researchers.

Discussion

We have introduced a Bio Community Cluster farm, BioCC, and its five major Biowiki web based services: BioCourse, BioPedia, BioSpecies, BioPeople, and BioCorea. Biowiki sites in BioCC form an organic information network whose contents are modifiable by anonymous users. While this enables a very fast and up-to-date knowledge exchange amongst internet communities, there can also be copyright problems that could cause originality disputes. To settle this possible issue, BioCC advocates a license scheme called Biolicense. Biolicense (*http://biolicense.org*) enables any human being and machine to openfreely share information and knowledge for a limitless number of purposes. It is a share- alike license that aims to protect biological information and knowledge from being legally monopolized by a small number of companies, classes, races, and economic groups in the world. The most important aspect of BioCC is that it is based on voluntary contribution. Users manage and maintain the information in BioCC. If users accept the underlying philosophy of BioCC, that research information should be freely exchangeable through an open-minded community, BioCC can become a valuable human heritage that should be transferred to future generations who will live in the era of 'personal omics' such as personal genomics.

Acknowledgements

This work was supported by M10407010001-06N0701-00110, and M10508040002-06N0804-00210 grant of MOST. JKK was supported by R01-2004-000-10172-0 (2005) grant of KOSEF. SSG would like to acknowledge all the supports of KOBIC in his previous period of stay and thanks his colleagues at KOBIC and OITEK, Inc.

References

Berners-Lee, T., Cailliau, R., Luotonen, A., Nielsen, H.F., and Secret, A. (1994) The World-Wide Web. *Communications of The ACM* 37, 76-82.

Cunningham, W. and Leuf, B. (2001). The Wiki Way Collaboration and Sharing on the Internet. Boston, MA: Addison-Wesley Professional.

Guest, D.G. (2003). Four futures for scientific and medical publishing. It's a wiki wiki world. *B.M.J.* 325, 1472-1475.

Maier, H., Dohr, S., Grote, k., o'keefe, S., Werner, T., Hrabe de Augelis, M., and Scheneider, R. (2005). Litminer and Wikigene; identifying problem-related key players of gene regulation using publication abstracts. *Nucleic. Acids Res.* 33, W779-W782.

Kai, W. (2006). Gene-function wiki would let biologists pool worldwide resources. *Nature* 439, 534.

Kevin, Y. (2006). Wiki ware could harness the internet for science. *Nature* 440, 278.

Stephen, C. (2006). Wiki and other ways to share learning online. *Nature* 442, 744.