

# **BIO**AUTOMATION, 2009, **13** (3), 97-110

## Topics of Bioengineering in Wikipedia

#### Vassia Atanassova

Centre of Biomedical Engineering – Bulgarian Academy of Sciences Acad. G. Bonchev Str., bl. 105, Sofia 1113, Bulgaria

E-mail: vassia.atanassova@gmail.com

Received: September 25, 2009 Accepted: October 5, 2009

Published: October 14, 2009

Abstract: The present report aims to give a snapshot of how topics from the field of bioengineering (bioinformatics, bioprocess systems, biomedical engineering, biotechnology, etc.) are currently covered in the free electronic encyclopedia Wikipedia. It also offers insights and information about what Wikipedia is, how it functions, how and when to cite Wikipedian articles, if necessary. Several external wikis, devoted to topics of bioengineering, are also listed and reviewed.

Keywords: Wikipedia, Bioengineering.

#### Introduction

Someone once joked that if something did not exist on Wikipedia, it did not exist at all. Of course, exaggerations and anecdotes of this kind normally rise around various concepts and ideas that gain unusual fame and momentum. But despite that being on Wikipedia is not an end in itself, it is getting increasingly harder with time to deny its growing role about informing the mankind and popularizing individuals, organizations, separate ideas and whole branches of knowledge and science.

The present paper aims to make a state-of-the-art report of how the scientific branch of bioengineering is presented in the free electronic encyclopedia Wikipedia, and more precisely: in the English Wikipedia, which is the largest and best visited one among its sister wikis, run by the Wikimedia Foundation, and in general. Along with this snapshot, some of the basics of Wikipedia will be presented with respect to its technological specificities, principles for content and code of conduct, reliability, as well as tips for citing Wikipedian content in scientific articles, if necessary.

### Main concepts of Wikipedia

Before commenting on Wikipedia, we should first discuss what a wiki is and how it functions. Wiki is a type of collaborative software that allows web pages to be created and collaboratively edited using a common web browser. Websites running such programs are themselves referred to as wikis. A wiki system is usually an application server that runs on one or more web servers. The content is stored in a file system, and changes to the content are stored in a relational database management system. The minimal requirements for launching a wiki are:

- a web server like Apache or IIS;
- PHP version 5.0 (or later);
- database server like MySQL 4.0 (or later) or PostgreSQL 8.1 (or later).



Some users find it helpful to install an additional software package such as phpMyAdmin (MySQL) or phpPgAdmin (Postgres) to help administer the database server, [5].

Wiki software allows the easy creation and editing of any number of interlinked webpages, using a simplified markup language or a WYSIWYG text editor, within the browser. Wikis are often used to create collaborative websites, to empower community websites, to collaborate in corporate intranets and knowledge management systems, or simply for personal note taking. Wikipedia is example of the first purpose. The English language Wikipedia has the largest user base among wikis on the World Wide Web and ranks in the top 10 among all websites in terms of traffic, [1, 13].

The subtitle of Wikipedia is The Free Encyclopedia and it often needs elucidation due to the multiple relevant meanings of the English word "free". First of all, Wikipedia is pay-free and advertisement-free. It is and the only non-profit website among top ten, maintained mainly by means of donation campaigns among the readers and editors.

On the other hand, Wikipedia is free with respect to the technical opportunity given to its readers to edit its content, which is the core of the wiki software's philosophy. It provides a transparent medium where any user may edit, and all edits and actions are logged under unique identifiers in the page history and the user's list of contributions, and associated with the respective username or IP-address, date and time. Thus, each edit remains documented and accessible over time and can be compared to any other version of the page, so that changes are easily traced and present an evidence of user's activities on the website.

The third and most notable aspect in which Wikipedia is free is its copyright policy. Wikipedia is published under the free content licenses of Creative Commons – Attribution – Share-alike (CC-BY-SA, [4]) and GNU Free Documentation License (GFDL, [3]). Both of these licenses attribute the licensed content with the following four freedoms of information (as formulated in the Definition of Free Cultural Works, [2], based on the definition of free software by the Free Software Foundation, [14]):

- the freedom to use the work and enjoy the benefits of using it;
- the freedom to study the work and to apply knowledge acquired from it;
- the freedom to make and redistribute copies, in whole or in part, of the information or expression:
- the freedom to make changes and improvements, and to distribute derivative works.

Users of freely licensed content are implicitly given these freedoms and need not to search for further explicit permissions from the content's author or copyright holder. However, users who wish to distribute copies or the original or modified content (freedoms 2. and 4.) are obliged to keep the following two requirements:

- to provide proper attribution of the content's author(s) and source, and
- to license the copies and derivative works under the same license.

The second requirement is called share-alike (in CC-BY-SA) or copyleft (in GFDL) and it means that what has once been pronounced free, remains free forever.



The encyclopedic content of Wikipedia generally consists of:

- encyclopedic articles,
- various forms of organization and navigation of the content, like:
  - o disambiguation pages and redirects;
  - categories:
  - o templates (navigational templates, infoboxes, stub templates, etc);
  - o lists;
  - o portals.
- various forms of enabling discussion, interaction and collaboration among the members of the community, like:
  - o articles' and users' talk pages, and
  - o wiki projects.

Articles in Wikipedia are the main object of development and discussion. With respect to their comprehensiveness, length and quality, they may vary from stubs to featured articles. Interwiki links connect articles in the English Wikipedia with their analogues in different languages as well as with galleries of photos, diagrams and other illustrative content. Despite their length, articles should be written with respect to three main content principles: Verifiability, Neutral point of view, and No original research, [11]. Some readers expect that articles have to present The Truth. In most cases articles do so and are even very close to perfection. However, there are also many controversial topics where no single truth can be outlined, and these are the cases when applying the leading content principles is the only warranty for a relatively good result. The reasons behind the formulation of these three main content principles are to be sought in the desired high quality standards against the technical capability given to virtually all readers to edit.

Articles' verifiability means that all data, statements, interpretations and conclusions in the articles should be derived from cited reliable and accessible published sources of information. Depending on the particular theme of the article, different sources of information are deemed reliable, but when it comes to articles on a scientific topic, such are the serious peer-reviewed scientific journals within the field, ideally those of high impact factor. Of course, the fact that a statement is published in a refereed journal does not make it automatically true. Even a welldesigned experiment or study can produce flawed results or fall victim to deliberate fraud.

The second major principle of the Wikipedia content is the so-called *Neutral point of view*. The neutral point of view is an approach of dealing with conflicting perspectives on a topic as evidenced by reliable sources. It requires that all majority and significant minority views must be presented fairly, in a disinterested tone, without bias, and in rough proportion to their prevalence within the source material. Article content should clearly describe, represent, and characterize disputes within topics, but should not endorse any particular point of view. Instead, articles should provide background on who believes what, and why, and on which points of view are more popular. Neutrality in scientific context is sometimes referred to as scientific consensus. Sources for this scientific consensus can be found in recent, authoritative review articles, textbooks, some forms of monographs and major up-to-date reference works such as medical dictionaries or scientific encyclopedias. However, special attention shall be paid to over-simplifications that are likely to be found in condensed dictionaries and encyclopedias.



The third core principle, No original research, is the one that mostly distinguishes Wikipedia articles from the true scientific publications, whose aim by definition is to announce new original research results and discoveries. On Wikipedia unpublished facts, arguments, speculation, and ideas are undesirable, as well as any unpublished analysis or synthesis of published material that serves to advance a position.

The MediaWiki software, which powers Wikipedia, provides various approaches to organizing and navigating the content, which here will be briefly described.

Disambiguation pages help navigation in cases when a given title may refer to two or more articles, while redirects facilitate finding a given article under two or more titles.

Categorization of articles is a mighty and flexible technique for organizing and browsing the wiki content. A category page provides access to exactly three levels of the category graph at a time: its own level, one level up (links to parent categories) and one level down (links to children categories, if any). More detailed rendering of the category graph, starting from a particular node, is possible using a special extension, available in Wikipedia. Categorization is done by putting the definition of the desired category (or categories, if more than one are relevant) within the body of the article. Thus, the articles call categories, and not the other way round. Categories only render the lists of pages and subcategories that call them, without being able to manipulate them directly. Pre-categorization occurs by editing the articles and subcategories, rather than editing the category itself, so it may become a problem when a large category has to be renamed. However, when it comes to building or restructuring the category graph, these are easy jobs using this very paradigm of categorization.

Templates in the MediaWiki software allow the inclusion of one and the same piece of information in multiple pages at a time, providing a centralized way to manage this content and automatically refresh it everywhere upon change. There are generally two kinds of templates for articles. Infoboxes are topic-specific templates which contain blank fields that require filling up with particular data about the subject. Navigational templates contain predefined, relatively stable sets of links to articles, lists of topics and/or categories which form relatively small semantic networks.

Lists are pages from the main namespace which differ from regular articles in their purpose and layout. They aim to present in a structured way sets of links to articles and/or categories. Lists are usually bulleted without any special formatting, but sometimes they are rendered in the form of a table, which is sortable according to its columns.

There are certain differences between lists, categories and templates with respect to the ways they facilitate the user navigation and understanding of the comprehensiveness of a topic. Even in cases when they apparently duplicate their functions, lists, categories and templates are to be used in synergy, rather than to replace each other.

Lists and navigational templates differ from categories in the way that they may contain "red links" to yet unwritten articles that were deemed relevant to a given topic or semantic network. Categories cannot contain these "red links": a red link means a non-existent page, while assigning a page to a category requires putting at least a definition of this category. Yet, pages that do not contain reasonable content to qualify for an article, but contain only a category call, are due to speedy deletion.



Templates also differ from lists in a way. A navigational template graphically represents a relatively small box placed in the bottom of all articles, linked within it. It functions to outline the major focal points of a topic and it may not aim to be fully comprehensive or detailed (although certain templates are). However, the volume of the template may not increase to extremities and since it affects multiple pages at a time it requires careful and rare editing. In comparison, editing a list affects only its own page and there are no limitations to the volume and level of detail which can be reached in presenting the topic.

The idea of the wiki portals is to help readers and/or editors navigate their way through Wikipedia topic areas through pages similar to the Main Page. In essence, portals are useful entry-points to Wikipedia content. Most portals contain the following for readers: a selected article and/or picture, links into the main category for the topic and possibly subcategories, Collaboration of the week, month, etc., general info about the subject, or links thereto, links to other portals and related WikiProjects (see below).

Apart from creating and improving encyclopedic content and facilitation its navigation, a good share of human and website resources are allocated to discussions and interactions. These are enabled by the talk pages that accompany pages from all namespaces (the main namespace, namespaces User:, Category:, File:, Image:, Template:, Wikipedia:, Mediawiki: and Portal:). Talk pages are mainly used for discussions by two or several editors, while the real group collaboration is done in the WikiProjects.

WikiProjects are pages in the community namespace Wikipedia:, which form collections of pages devoted to the management of a specific topic or family of topics within Wikipedia; and, simultaneously, enables a group of editors to collaborate on encyclopedic work. It is not a place to write encyclopedia articles directly, but a resource to help coordinate and organize the writing and editing of those articles. The projects are used to plan, schedule, discuss and report the work on certain aspects of the presented topic. Some WikiProjects occur spontaneously among seasoned wikipedians, while others are designed as a part of the training process in schools and universities or aim at involving new editors to the pool. Since 2006, there are more than a hundred initiatives in high school and universities around the world, which have resulted in the recruitment of thousands of students to contribute to Wikipedia, thus presenting their knowledge on a topic, relevant to their studies, under the guidance and supervision of their teachers and professors, [12].

### Wikipedian Coverage of Bioengineering Topics

The above presentation of the main concepts of Wikipedia was necessary for the survey of how topics from the field of bioengineering are covered in the free encyclopedia. The scope of the survey was intentionally limited to the English Wikipedia that is the first and the largest wiki in the family of the Wikimedia Foundation. It is noteworthy that the present survey does not aim to assess the quality and exhaustiveness of the Wikipedian content on bioengineering, but only provide a snapshot of this content as of September 2009.

The most general category, Bioengineering, is a child of the following categories: Engineering disciplines, Biology, Health sciences, Medical research, and Medical treatments. Among its subcategories, on different levels of hierarchy following categories can be found: Artificial Bioinformatics, Bioluminescence, engineering, organs, Biomedical Biopharmaceuticals, Bioreactors, Brain-computer Bioremediation, Biotechnology,



interfacing, Cell imaging, Cloning, Genetic engineering, Microarrays, Molecular genetics, Neural engineering, Neuroprosthetics, Prosthetics, Synthetic biology and others less relevant. Special attention is also paid to auxiliary categories like Bioengineers and Biotechnologists, as well as categories Biotechnology companies, Biotechnology organizations, Biotechnology law, Biotechnology literature, Biotechnology products.

Starting from the topmost node, passing through its directly related subcategories from the field of bioengineering and removing the multiple occurrences due to overcategorization, more than 2000 distinct Wikipedian articles were found.

Various lists of topics within the field were also detected, using the Category Intersect tool on the toolserver of Wikimedia. A part of them is the following:

- Bioinformatics workflow management systems;
- List of bioinformatics companies;
- List of bioinformatics journals;
- *List of biotechnology companies*;
- List of phylogenetic tree visualization software;
- List of phylogenetics software;
- *List of protein structure prediction software*;
- List of quantum chemistry and solid state physics software;
- *List of RNA structure prediction software*;
- List of sequence alignment software;
- List of sequenced eukaryotic genomes;
- List of software for molecular mechanics modeling;
- *List of Y-DNA single-nucleotide polymorphisms*;
- Software for protein model error verification;
- Software for protein structure visualization, and others.

Within the scope of the present survey several wiki projects were found relevant. These are: WikiProject Chemical and Bio Engineering with 33 participants, WikiProject Mathematical and Computational Biology with 13 participants, and WikiProject Molecular and Cellular Biology with 265 participants. These sub-communities have elaborated on two wiki portals that may be of particular interest: Portal:Biotechnology and Portal:Molecular and Cellular Biology. Figures 1(a, b) and 2(a, b, c) present screenshots of both portals, located on the webpages

http://en.wikipedia.org/wiki/Portal:Biotechnology and http://en.wikipedia.org/wiki/Portal:Molecular\_and\_Cellular\_Biology.

There were held several university projects, relevant to the field of bioengineering, [12]:

- An undergraduate project of the School of Chemical and Biomolecular Engineering in the University of Sydney, as of March 2008.
- A project on junior/senior level engineering course of 9 students from the Cornell University in Fall 2008 aimed at adding content in the area of bioseparations (purification of proteins and other natural products). Among topics selected for coverage are centrifugation, fast protein liquid chromatography, and apheresis.
- An end of term project for a senior level engineering course on downstream processing in Cornell University in Fall 2006. As a result, eleven students created or expanded articles on unit operations and equipment for production of purified biosynthetic pharmaceuticals and refined food ingredients.

A project of Yale University in Fall 2006 for the students in "History of Modern Science in Society" lead to the creation, among all, of an article on history of biotechnology.

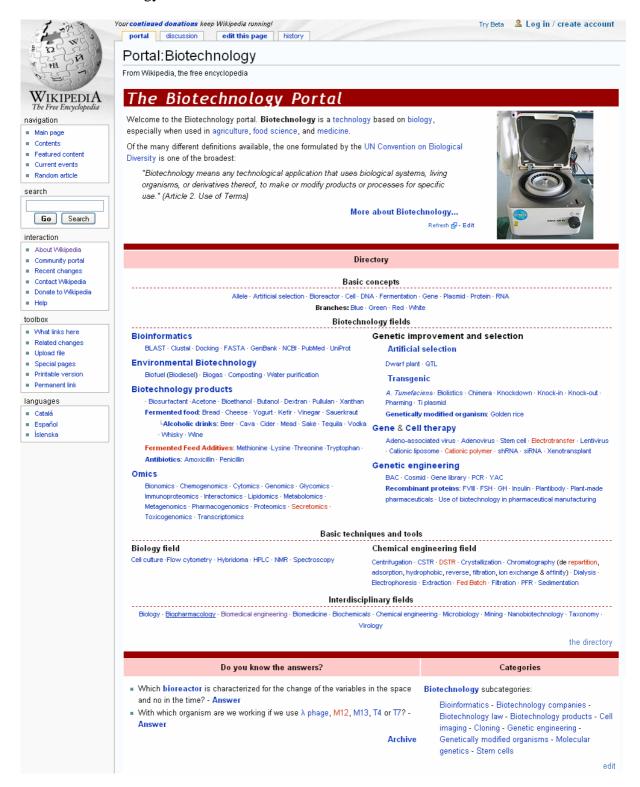


Fig. 1(a) Portal "Biotechnology" on the English Wikipedia

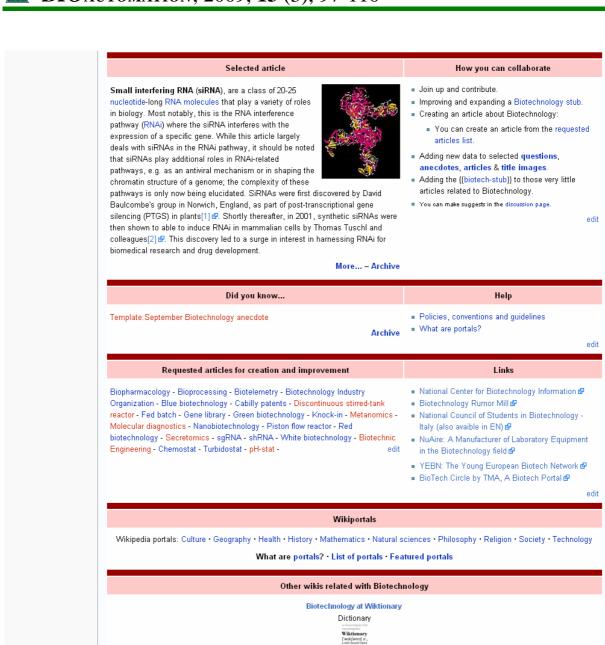


Fig. 1(b) Portal "Biotechnology" on the English Wikipedia

apply. See Terms of Use for details

Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may

Categories: Biotechnology | Portals under construction

Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.

This page was last modified on 14 May 2009 at 14:14.

WIKIMEDIA

edit

[ Powered Media



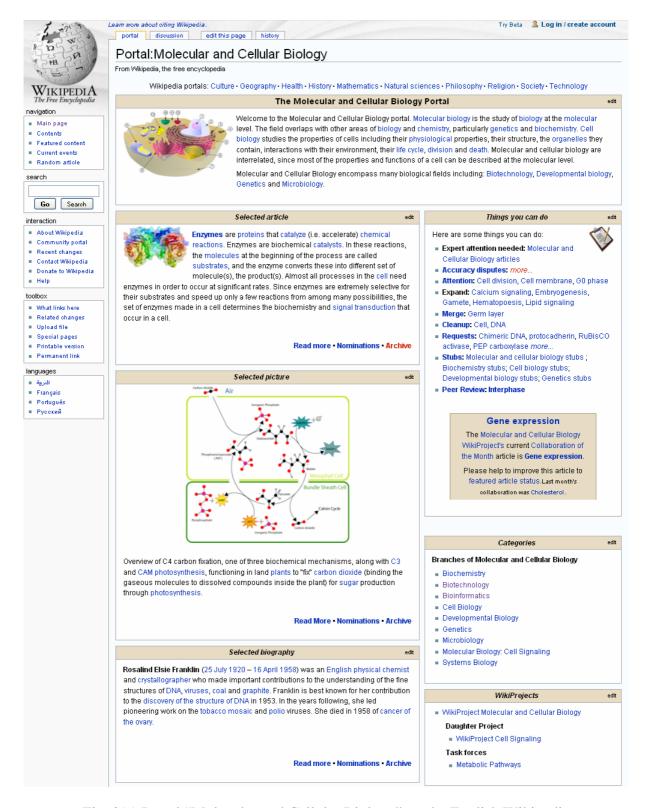


Fig. 2(a) Portal "Molecular and Cellular Biology" on the English Wikipedia



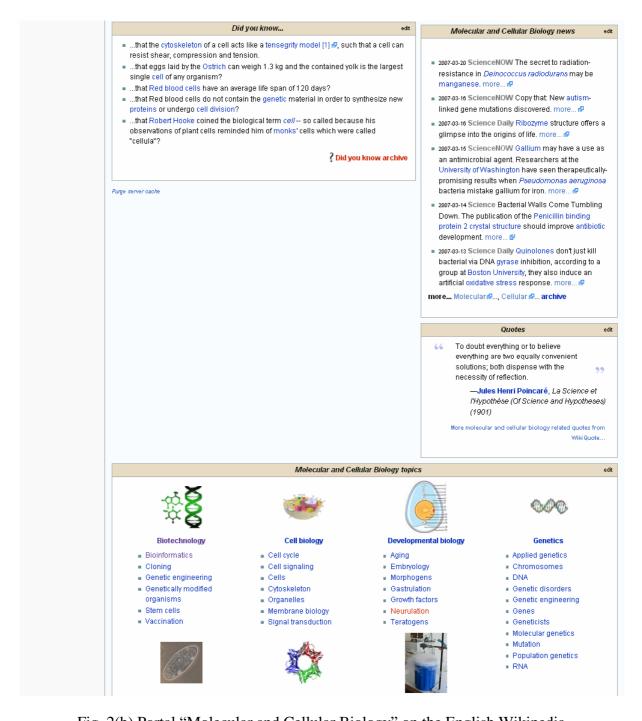


Fig. 2(b) Portal "Molecular and Cellular Biology" on the English Wikipedia

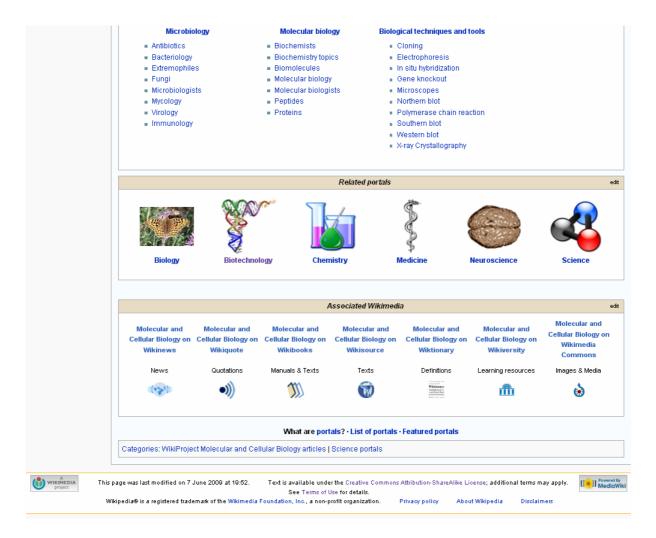


Fig. 2(c) Portal "Molecular and Cellular Biology" on the English Wikipedia

Apart from the bioengineering content of Wikipedia, there exist other websites, powered by the wiki technology, which are devoted to topics in the field. Of specific interest may be the following projects, chosen here due to their currently active state:

- Medpedia a collaborative project launched in February 2009 with the aim to create an open access online medical wiki encyclopedia in association with Harvard Medical School, Stanford School of Medicine, Berkeley School of Public Health, University of Michigan Medical School, the U.K. National Health Service (NHS), as well as other contributors. Content is licensed under the GFDL and runs on MediaWiki software. Medpedia is comprised of three primary components: a collaborative encyclopedia (knowledge base), a Network & Directory for health professionals and organizations and a community of interest where medical professionals and non-professionals come together to discuss topics of interest, [7]. The website is located on <a href="http://www.medpedia.com/">http://www.medpedia.com/</a>.
- *SNPedia* (pronounced *snipedia*) a semantic wiki on bioinformatics that serves as a database of single nucleotide polymorphisms (SNPs). Each article on a SNP provides a short description, links to scientific articles and personal genomics websites, as well as microarray information about that SNP. In September 2009, the website claimed to have 7,938 SNPs in their database. An associated freeware computer program called Promethease, also developed by the SNPedia team, allows users to compare personal



genetics results against the SNPedia database, generating a report with information about a person's attributes, such as propensity to diseases, based on the presence of specific SNP's within their genome, [9]. The website is located on: http://www.snpedia.com.

- OpenWetWare a wiki whose mission is "to support open research, education, publication, and discussion in biological sciences and engineering". It was created by graduate students at MIT on April 20, 2005. Initially, it served as a private lab wiki, but was opened up to allow any lab to join on June 22, 2005. As of April 6, 2007 the site hosted 100 research laboratories from over 40 institutions, including Boston University, Brown University, Caltech, Cambridge Research Institute, CNRS, Duke University, and many others. In addition to laboratories, a number of scientific communities are based on the site, including synthetic biology, Mimulus, and the BioBricks Foundation. OpenWetWare runs on MediaWiki software on Linux servers. All content is available under the GNU Free Documentation License and the Creative Commons Attribution ShareAlike license, [8]. The website is located on: http://openwetware.org.
- Wikigenes is a WikiGenes is a collaborative knowledge resource for the life sciences, which is based on the general wiki idea but employs specifically developed technology to serve as a rigorous scientific tool. The rationale behind WikiGenes is to provide a platform for the scientific community to collect, communicate and evaluate knowledge about genes, chemicals, diseases and other biomedical concepts in a bottom-up process, [6]. The website is located on: http://wikigenes.org/.

### Citing Wikipedia

As with any source, especially one of unknown authorship, users should be wary and independently verify the accuracy of Wikipedia information, if possible. For many purposes, but particularly in academic publications, Wikipedia may not be an acceptable source. This is especially true when it is used uncorroborated. However, much of the content on Wikipedia is itself referenced, so an alternative is to cite the reliable source rather than the article itself.

Special care should be taken when using Wikipedia as a source for research projects. Normal academic usage of Wikipedia and other encyclopedias is for getting the general facts of a problem and to gather keywords, references and bibliographical pointers, but not as a source in itself. Wikipedia is a wiki, which means that anyone in the world can edit an article, deleting accurate information or adding false information, which the reader may not be aware of.

A wiki is an unusual medium, and as such does not conform well to the usual citation formats for books and journals. The exact format shall depend upon the citation guide of the respective book or journal, but there are several useful considerations:

There is a special tool assisting the citation of Wikipedian pages. On the left of every article, there is a "Cite this article" link. Clicking it brings the user to a listing of relevant information, as well as automatically generated citations in several bibliographic styles: APA style, MLA style, MHRA style, Chicago style, CBE/CSE style, Bluebook style, Bluebook: Harvard JOLT style, AMA style, and style of a BibTeX entry.



- Wikipedia is collaboratively written. This means that no particular author or authors may be listed when citing a particular Wikipedian article. However, if the list of authors is needed, it can be taken from the page history. Authors are listed only by IP address or username; so it is usually impossible to verify and even guess their identities.
- Citation of a Wikipedian article should normally include the article title, the publisher Wikipedia, The Free Encyclopedia, as well as the date of access (usually, this is the date of the latest version of the page). Every article should be a separate citation.
- Most citation styles will likely require the full article URL. It can be derived by clicking on the "Permanent link" link in the toolbox on the page's left sidebar. This lets the URL include the unique identifier of the visited page version, which provides the most accurate referencing to this otherwise dynamically changing content, [10].

### **Conclusion**

Apart from giving an idea of what Wikipedia is and how it covers the area of bioengineering, this article aims to reach for new potential editors among the readers. There are many good reasons for researchers and university professors to personally contribute to The Free Encyclopedia or involve their students to do so: to popularize their scientific field and published results among the general audience and other specialists, to practice teamwork and master the delivery of neutral and well referenced encyclopedic texts, as well as share Wikipedia's vision for a world in which every single person is given free access to the sum of all human knowledge.

#### References

- 1. Alexa Top 500- Global Sites, http://www.alexa.com/topsites (as of 25-09-2009)
- 2. Definition of Free Cultural Works, http://freedomdefined.org/index.php?title= Definition&oldid=5437
- 3. GNU Free Documentation License, <a href="http://www.gnu.org/copyleft/fdl.html">http://www.gnu.org/copyleft/fdl.html</a>
- 4. Legal code of Creative Commons Attribution Share-alike License, http://creativecommons.org/licenses/by-sa/3.0/legalcode
- 5. Mediawiki contributors, "Installation", Mediawiki.org, http://www.mediawiki.org/w/ index.php?title=Installation&oldid=273771 (as of 31-08-2009).
- 6. Robert Hoffmann, "A wiki for the life sciences where authorship matters", Nature Genetics 40, 1047-1051 (2008) http://www.wikigenes.org/e/art/e/1.html?vs=6 (as of 27-08-2008).
- 7. Wikipedia contributors, "Medpedia," Wikipedia, The Free Encyclopedia, http://en.wikipedia.org/w/index.php?title=Medpedia&oldid=314202369 (as of 15-09-
- 8. Wikipedia contributors, "OpenWetWare," Wikipedia, The Free Encyclopedia, http://en.wikipedia.org/w/index.php?title=OpenWetWare&oldid=314103028 (as of 15-09-2009).
- 9. Wikipedia contributors, "SNPedia," Wikipedia, The Free Encyclopedia, http://en.wikipedia.org/w/index.php?title=SNPedia&oldid=312520878 (as of 08-09-
- 10. Wikipedia contributors, "Wikipedia: Citing Wikipedia", Wikipedia, The Free Encyclopedia, http://en.wikipedia.org/w/index.php?title=Wikipedia:Citing\_ Wikipedia&oldid=310731024 (as of 29-08-2009)



- 11. Wikipedia contributors, "Wikipedia: Five pillars", Wikipedia, The Free Encyclopedia, http://en.wikipedia.org/w/index.php?title=Wikipedia:Five\_pillars&oldid=316244555 (as of 26-09-2009)
- 12. Wikipedia contributors, "Wikipedia:School and university projects," Wikipedia, The Free Encyclopedia, http://en.wikipedia.org/w/index.php?title=Wikipedia:School\_and\_ university\_projects&oldid=315833950 (as of 24-09-2009)
- 13. Wikistats by S23, http://s23.org/wikistats/largest\_html.php?sort=users\_desc&th= 8000&lines=500 (as of 25-09-2009).
- 14. The Free Software Definition, http://www.fsf.org/licensing/essays/free-sw.html

### Vassia Atanassova, Ph.D. student

E-mail: vassia.atanassova@gmail.com



Vassia Atanassova has graduated Informatics in Sofia University in December 2004. She works in the Centre of Biomedical Engineering -BAS, and is a PhD student in the Institute of Information Technologies – BAS since June 2009. She is an editor of Bulgarian Wikipedia since June 2006, administrator and active promoter of free content and free software.