

Learning with the Web 2.0: The Encyclopedia of Life

Anwar Us Saaed¹, Alexander Stocker², Patrick Hoefler³, Klaus Tochtermann⁴

Knowledge Management Institute, Graz University of Technology¹⁴; Institute for Networked Media, Joanneum Research⁴; Know-Center Graz²³⁴

Key words: *Web 2.0, Encyclopedia of Life, Wikipedia, learning, e-learning, knowledge transfer*

Abstract:

In the 1990s, the World Wide Web has revolutionized the way in which knowledge seekers satisfy their thirst for knowledge. During the recent years, a technological and social paradigm shift called Web 2.0 attracted a lot of attention in the Internet community and is considered a major evolution of the web. New online applications make it easier for individuals to learn, using the vast information provided by the Internet and the collective intelligence of its users. The Encyclopedia of Life, a global repository for all kinds of information related to life on earth, builds upon the vision of Wikipedia and enhances it with Web 2.0 technologies and a concept for assuring high quality content. After a brief review of the development of the web as a learning resource, the authors of this paper present the concept of the Encyclopedia of Life as a knowledge pool for the domain of biology and compare it to the Wikipedia.

1 Introduction

Technology has played a major role for learning, knowledge creation and knowledge transfer in the modern age — beginning with Gutenberg's printing technology and more recently leading to digital publishing on the web and mass production of user-generated content as a result of the Web 2.0 phenomenon. New social web tools and applications enable users to be the masters of the information. The unbound number of content creators have spurred a new age of information, leading to a revolutionized experience of learning. In the Web 2.0 era of collaborative technologies, the mass production of blogs, wikis, podcasts and websites brings in mountains of partly redundant information and demands a great effort to develop an accumulative understanding on a topic. Therefore, a need for the encyclopedic aggregation of knowledge has never been felt stronger than in the Web 2.0 age, even in the presence of 'googling' technologies. Wikipedia [16] was the first to take up this challenge and with the enthusiasm of 'social text' gathered more than 5,5 million entries up to now (September 2007), achieving a milestone in this regard.

Wikipedia serves well for informal learning and building a basic understanding of a given topic, but for scientific purposes, it remains unsatisfactory [5]. Furthermore, the need for an authentic, authoritative and comprehensive source of knowledge on a particular topic has not been satisfied. Although many initiatives like Citizendum [1] or Scholarpedia [13] have tried to fill this gap, they usually only provide an overview and do not cover a field of knowledge in depth.

The desire to understand life forms on our planet is not new. The success of the Genome project, 'one of the most significant achievements of modern science' [9] and the technological advancement in biology and informatics provide the foundation for 'a leap for all life' [4] — the Encyclopedia of Life (EOL) [2]. It is envisioned as the first major

encyclopedia of the Web 2.0 that will cover the breadth and depth of authentic and comprehensive information as ‘a macroscope for biodiversity and an entry point into virtually all of biological knowledge’ [11]. It also aims to ‘combine the authority of a traditional print behemoth with the collaborative spirit of the Web’s user-created Wikipedia’ [8] to create a separate web page for each species on earth.

In section 2, this paper introduces the concept of Web 2.0 and learning while collaborating on the web. Section 3 introduces the Encyclopedia of Life (EOL) as a global knowledge repository for the domain of biology, developed with a user-centered approach using Web 2.0 technologies. Section 4 compares the concept of the EOL with the successful online encyclopedia Wikipedia. Section 5 concludes with a summary and an outlook for future developments and research.

2 Effects of the Web 2.0 on Learning

The term Web 2.0, which has attracted a lot of attention in the Internet world, has been coined to describe the changes that the Web is currently going through [10]. Most of them are caused by the vast growth of the web together with the rise of new collaborative technologies, reaching out for a richer user experience. Web 2.0 is, at the same time, a social phenomenon, causing users to interweave their communication and interaction processes with the web. Users have continually begun to assemble in new types of online communities which are emerging all over the web [7], accompanied by changing their traditional role from mainly using the Internet as a source of information to actively participating in the content creation process. The social phenomenon is enabled by the technical revolution, where new rising technologies including content syndication, semantic annotation and richer user interfaces are tempting social interaction, thus resulting in the emergence of new types of collaborative knowledge structures on the web. Social entities [15] act as a resource for collaborative knowledge where the web in total can be seen as a collaborative learning environment.

Entry barriers of using the web as a medium for learning have been reduced mainly due to, amongst others, the radical simplification of interactive user interfaces and easy access to huge pools of knowledge. This has changed the way the individual knowledge workers are gaining their context-specific knowledge. Google.com, not only because of its simple user interface, has been one of the precursors of an information aggregator, allowing the learner to search quickly and easily for proper learning resources.

Due to the nature of the web, theme specific niche communities emerge. The voluntarily conducted knowledge transfer between members in the respective communities is embedded in the social structures, enhancing the communities. The Web 2.0 provides, amongst other features, the possibility of instant communication without the requirement of any additional tools or plug-ins. Technological advances like folksonomies and semantic annotation facilitate the process of finding users with similar interests or problems. Because of the socializing nature of these Web 2.0 knowledge communities, the knowledge seeker will get an answer to his or her questions straightforward, compared to monotonous e-learning platforms where the motivation of the users to contribute and participate may be much lower. Web 2.0 furthermore enhances the learning skills while practicing learning by doing, e.g. writing a blog about a specific topic of interest and getting feedback from the audience, thereby increasing the knowledge of the blog author and all of its readers.

‘Traditional’ e-learning systems often lack the consideration of social factors like trust and motivation [6]. Rollet [12] gave an overview on the implications of Web 2.0 to knowledge transfer and investigated the potential of lightweight Web 2.0 applications like blogs and wikis for the purpose of facilitating self-directed collaborative responsible learning in educational settings.

3 The Encyclopedia of Life (EOL)

3.1 Introduction

In this section, the authors of this paper give an overview of the Encyclopedia of Life (EOL) [2], a comprehensive knowledge repository for the domain of life sciences. Concept, features, stakeholders and underlying technologies are discussed.

At this time, the development of the EOL is still in progress, and the first species pages will be made accessible in the mid of 2008. Documents concerning the design, the work-plan, technological issues and vision along with mock-ups of the user interface are available to the public and are the main source of information for the following analysis.

The vision of the EOL is not a new one: Already in the 1990s, Daniel Janzen from the University of Pennsylvania was among the first to address species pages. More than 10 years later, E.O Wilson articulated Janzen's idea in his essay "The Encyclopedia of Life" [18] and became one of the leading proponents of the EOL.

As stated, the goal of the EOL is to serve as an online reference source and database for each and every of the 1.8 million species that are known and named today, and for those who are still to be discovered. The EOL aims to be used as a teaching and learning tool for scientists, students, educators and everyone interested in the domain of biology. Students and educators will find a variety of tools to use the already available information and feed their own knowledge back into the EOL.

The emergence of the EOL is mainly caused by three factors: First, the necessary consortium has been brought together, inspired by E.O. Wilson's vision at the TED conference in March 2007 [19]. Second, due to the rapid deterioration of the ecological system on earth and the environment resulting in the extinction of countless species, there is a strong need for a repository to document all kinds of life. This repository could help preserving life forms before they are gone forever as there is no comprehensive database of every species available today. Third, in the recent years, crucial tools like semantic technologies and wiki-style editing have proven mature enough to be used on a grand scale.

By harnessing the work of scientists from all over the world, the EOL aims to become the global resource for information regarding all life on earth. A comparable knowledge pool has never been available to the scientific community or society before. Today, there are many information resources both online and offline — however, even the smartest searcher is exposed to information overload resulting in lost time and energy. The creation of the EOL is a collaborative effort, tens of thousands of experts from all over the world are responsible for its constructing, led by a steering committee of experts from renowned institutions in the field of biology such as Harvard University, Smithsonian Institute and Biodiversity Heritage Consortium.

The Encyclopedia of Life continues the historical practice of sharing knowledge to advance scientific innovation and learning and is envisioned as a computer-based, ever-expanding unified and structured chronicle of all life forms.

The content in the EOL is brought together from a wide variety of sources and will be authenticated by scientists and biologists to provide a quality-assured information resource. The EOL and its board will work with scientists all around the globe, securing that the people who are involved are experts in their respective field.

To ensure that information on the EOL stays current, regular and continued contributions are necessary. Software tools are created to mine the scientific literature. The scientific community will use EOL-developed tools for their scientific purposes to easily submit updates.

At the moment, the EOL is working on three major challenges: First, accommodation of the diverse demands of the different stake-holders involved needs a lot of conceptual work. Second, the realization of a user-friendly tool for all audiences needs a lot of effort on the part

of the technologists. Finally, the intellectual property rights issues are always a big challenge for technology-based knowledge sharing and will have to be resolved.

3.2 Key technologies within the EOL

The EOL has planned a very flexible and end-user focused information architecture [11] that can be perceived in three major steps: aggregation, authentication and atomization.

Knowledge about many species is already available in pieces, in different forms and from different sources. The aggregation will bring all the pieces from libraries and websites together to establish a more structured and comprehensive repository. The major technologies used for the aggregation will be APIs, semantic tools for editing and annotating content, and scanning and OCR technologies for digitizing analog content. These technologies will be customized in the form of ‘Taxonomic Intelligence’ [11] and GUIDs to resolve problems with synonyms and homonyms using conciliation and disambiguation techniques. The EOL intends to add semantic intelligence to biology for metadata standardization of the content, so that information can be identified, retrieved and cached as required, distributed and made accessible through an index. The ‘Taxonomic Name Recognition’ (TNR) tools use name discovery and name recognition algorithms for the indexing systems and gather previously undiscovered names for NameBank (a registry for all the recorded taxon names), holding the global name repository, and facilitating dynamic taxonomic navigation of RSS-Feed, Databases or web-pages. The whole structure of the EOL is linked by the species name which is the only field common to virtually all biological databases.

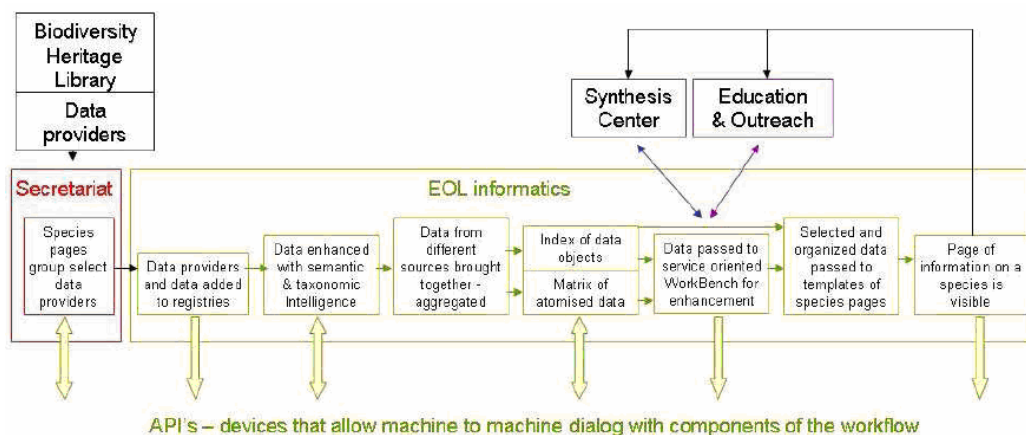


Figure 1: Assembling a species [11]

Content is aggregated from a wide variety of resources. The content gathered from different resources receives authentication from scientists, ensuring the learner have authoritative information. EOL will secure the involvement of the scientists and scientific institutions that are established experts on each species. The EOL will provide editing tools and a semantic wiki environment to engage the scientists around the globe for maturing of the content.

The EOL provides new ways of mining, visualizing and interpreting data using workbench tools and modules which is termed as atomization of gathered data fulfilling the needs of customized views. The EOL provides customization of the data in three different ways: First, the web interface of the EOL provides static fundamental information about the species. The dynamic part of the content presentation changes according to the level of the learner reaching from novice to expert. These content components are automatically obtained and structured by the machine using a big set of information from a wiki-type environment. Second, the myEOL customization can use workbench tools for the custom view of the information on a species along with some widgets, bookmarking and tagging tools. Third, the workbench tools

as the main source of atomization are developed according to the open source ideology using the community aiming to develop the consumer driven products for the future use of the information.

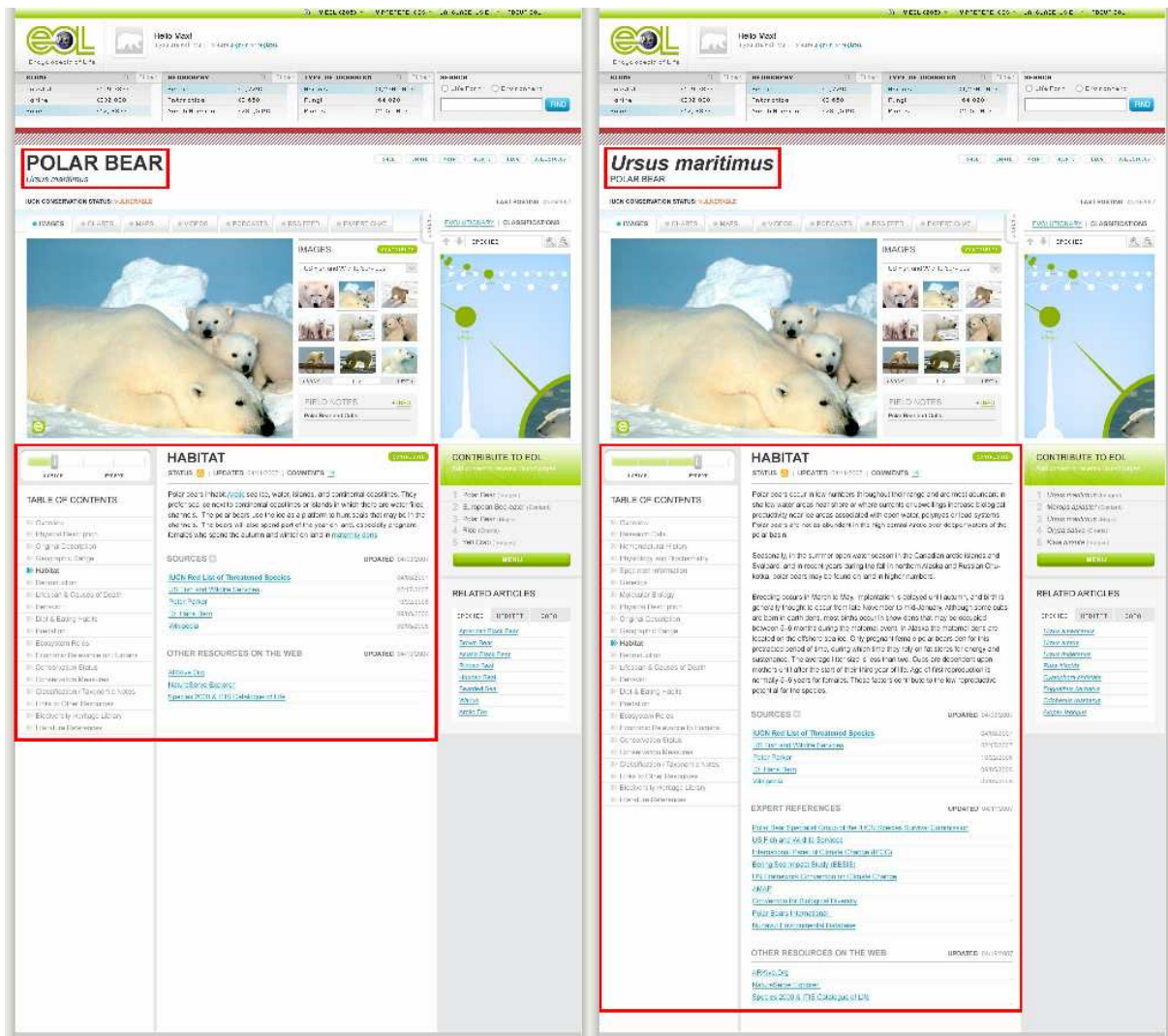


Figure 2: Novice and expert view [2]

3.3 Learning dimension of the EOL

The EOL presents the content and the multimedia data set in a way to engage learners from school level to the highest level of expertise achieved by scientists in a particular topic. Usability combined with latest technology ensure a high motivation of the learner to work with the EOL.

When using the Web as a learning medium, learners are facing the problem, that content which is appropriate for his or her particular skill level, is hard to retrieve. When the skill level of the learner increases, he or she must search for new appropriate learning resources on the particular topic. Wikipedia holds scientific articles which are under most conditions not understandable by the ordinary learner [5].

When using the EOL as a learning resource, the learner may adjust the content presented to him or her according to his or her skill level. The learner can benefit from a comprehensive learning resource on the topic of quality, making further searches for content obsolete. Content in the EOL is valid from a scientific point of view, allowing the learner to use it (if permitted) without scrutinising source or the learning material itself.

Interfaces for mobile devices allow the usage of the EOL in many situations including field studies, which are especially important for biologists. Learners may use the existing web-services or even create their own web-services, thereby enjoying the benefits of EOL's openness. By doing this, a greater impact on informal learning is achieved.

Schools and Universities enrich their courses and seminars by using Learning Management Systems (LMS) conducting blended learning. Most of the data in the EOL can be integrated in these learning environments, combining the functionalities and learning concepts underlying the Learning Management Systems with the ever fresh content in the EOL.

4 Comparison: Wikipedia vs EOL

In this section, the concept of the EOL will be compared with the concept of the Wikipedia. Wikipedia is chosen mainly because of two reasons: It is freely available and uses a similar authoring environment. Although both EOL and Wikipedia claim to be encyclopedias, they strongly differ in their goals. In general Wikipedia aims to build a widespread base of knowledge, in contrast EOL focuses to gather all the knowledge in the field of biology, creating a repository of the expert knowledge. As a result, articles in Wikipedia are numerous covering the breath of knowledge, but most of the time missing a detailed level, while EOL focuses on a particular topic, hence articles are expected to be on a consistently detailed level covering the depth in that topic. The comparison focuses on the three aspects content, stakeholder and technologies.

Wikipedia is a grown encyclopedia, addressing the phenomenon of mass authoring to the area of content creation in a wiki environment. Everybody may contribute to any subject in the Wikipedia regardless of his knowledge in the particular field. Wikipedia is suitable for providing an overview of a topic of interest towards a knowledge-seeker who can be anybody, including scientists. However, the usage of the content of the Wikipedia for scientific purposes is very limited, because of the lack of validity from the scientific community. In Wikipedia, plain text is dominating, multimedia content is scarce. Due to the collaborative nature of content creation, quality of content in Wikipedia is inconsistent and can easily be vandalized or falsified. Wikipedia lacks a workflow for quality assurance. A wiki-based discussion forum is aimed to support the collaboration of the authors and changes may be (but do not necessarily have to be) discussed there before they are conducted. Wikipedia also lacks personalization features and a bulk of content has to be browsed to find the relevant information on a topic.

Contrary to Wikipedia, the EOL uses a well defined workflow for information structuring and validation of content. Furthermore, EOL is enriched with personalization features to facilitate end-users to organize the content in the form they like it. For scientists, motivation of publishing within the EOL is high because, unlike Wikipedia, EOL holds the potential for reputation for the content creator. Due to the peer reviewed nature, EOL may even become a major platform for scientific publishing in biology in future.

The EOL incorporates pre-authenticated content as well as fresh content, which has to be peer reviewed by scientists, before being accessible to the public. The peer review is a formal authentication process, conducted by scientists, who are experts in the respective species. The pre-authenticated content is drawn from data-providers consisting of well-established research organizations from all over the world. Contributors for the fresh content may range from scientists to anybody with an interest in the domain of biology and biodiversity. Content in EOL will differ from content in Wikipedia regarding multimedia enrichments including images, audio and video.

Both Wikipedia and EOL use a wiki-like environment for the creation and usage of the content. Wikipedia supports the collaborative content creation with technologies like discussion wikis for each article and a revision control to counteract vandalism. Wikipedia offers no tools for reusing its content in different environments, except a simple option to download the whole Wikipedia in a huge file. Wikipedia lacks in current technologies including Tagging, Ajax or semantic ones. The Semantic Media Wiki [14] tries to enhance the Media Wiki, which is the underlying wiki for the Wikipedia, with Semantics.

The set of technologies in the EOL is very broad and has been introduced partly in chapter three. The EOL presents itself in an aesthetically pleasing way, offering vast multimedia support for the learner, out-rivaling the Wikipedia in the way the information is presented. The EOL is based on an interactive wiki-like environment. In the front-end the content elements are dynamically structured depending on the knowledge level of the learner by using a skill slider to select the expertise level. The pre presented figure 1 shows that the content in the novice level is more compact and easier to understand than in the expert level. When the slider is moved, both the available subtopics and the content of the article itself change according to the new skill level. News-feeds (RSS), podcasts and expert chats are provided to build a better understanding and up to date information on the topics of interest e.g. the latest scientific publications on a particular species. The EOL allows personalization of the content regarding the special needs of the learner by using bookmarking, tagging and widgets.

The EOL search is different to common search tools in the web, providing fine tuned semantic search mechanisms to cater for large and diverse set of end users. Due to the semantic algorithms, based on the underlying biological taxonomy, search is smarter and more relevant search results are retrieved. As an example, if one searches for the term 'habitat polar bear' the search result presented will be the corresponding content on the habitat of the polar bear. A taxonomic map visualized as a graph will show links between the polar bear and its related species.

Different to Wikipedia, the content cannot only be dumped to a file, but EOL provides sophisticated tools for reuse and mash-up of content. Based on the EOL content, modules can be developed allowing interested parties like research facilities or learning institutions to customize the interfaces or to conduct data mining according to their respective needs.

5 Conclusion

The EOL as described and in its status-quo is well planned and indeed revolutionary, using latest web technologies. Services will be available on a global basis to all the researchers and the learners using mobile and desktop devices. Learners may profit from a comprehensive knowledge pool in biology. To make the content of the EOL freely available to the learners, will be a crucial milestone. Schools and universities along with research institutions may use the EOL as a future platform for their projects fostering collaborative learning, hence not only using the content but also enhancing and expanding it. Expert-chat and data-mining facilities will be useful features in this regard.

The authors of this paper cannot propose, whether the EOL will be a success, or not. Nevertheless, a key success-factor for wiki-like environments is to reach the critical mass of both authors and articles. The EOL has planned to overcome this obstacle by focusing on an organized way of engaging the necessary stakeholders as data providers and authenticators. From the perspective of the authors, other knowledge repositories for authenticated content

did not succeed because they simply tried to copy the success patterns of Wikipedia without being innovative enough to attract a critical mass of authors. The EOL, being well focused, has engaged an expert-community from the beginning to achieve a critical mass of authors, offering the latest technologies to serve to the purpose of learning.

The authors of this paper furthermore assume that the popularity of Wikipedia as a whole will not decrease because of the EOL. However, an interesting topic will be the interplay between the EOL on the one side and the biological part of the Wikipedia and Wikispecies [17] on the other.

References:

- [1] Citizendium, <http://www.citizendium.org>, accessed 2007-08-22.
- [2] Encyclopedia of Life, <http://www.eol.org>, accessed 2007-08-22.
- [3] Encyclopedia of Life, Frequently Asked Questions, <http://www.eol.org/faqs.html>, accessed 2007-08-22.
- [4] Encyclopedia of Life, Press Release, http://www.eol.org/press_release.html, accessed 2007-08-22.
- [5] Epidemix Blog, Why does Wikipedia Suck on Science, <http://epidemix.org/blog/?p=72>, accessed 2007-08-22.
- [6] Fisman, R.; Khanna, T., Is trust a historical residue? Information flows and trust levels, in: Journal of Economic Behaviour and Organization, 1999.
- [7] Kolbitsch, J.; Maurer, H., The Transformation of the Web: How Emerging Communities Shape the Information we Consume, in: Journal of Universal Computer Science, 2006.
- [8] Science News, Extreme Encyclopedia: Every living thing will get its own page. <http://www.sciencenews.org/articles/20070512/fob7.asp>, accessed on 2007-08-22.
- [9] National Partnership for Advanced Computational Infrastructure: Archives, <http://www.npaci.edu/envision/v19.1/encyclopedia.html>, accessed on 2007-08-22.
- [10] O'Reilly, T., Design Patterns and Business Models for the Next Generation of Software, <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>, accessed 2007-08-22.
- [11] Patterson, D.J., EOL Biodiversity Informatics: Designing for a Living Encyclopedia of Life, eolinformatics.mbl.edu/Documents/WorkPlan.pdf, accessed 2007-08-22.
- [12] Rollett, H.; Lux, M.; Strohmaier, M.; Dösinger, G.; Tochtermann, K.; The Web 2.0 way of learning with technologies, in: Int. Journal of Learning Technology, 2005.
- [13] Scholarpedia, www.scholarpedia.org, accessed 2007-08-22.
- [14] Semantic Media Wiki, http://ontoworld.org/wiki/Semantic_MediaWiki, accessed 2007-08-22.
- [15] Stocker, A.; Dösinger, G.; Us Saed, A.; Wagner, C., The Three Pillars of Coporate Web 2.0: A Model for Definition, in: Proceedings of Triple-I'07, 2007.
- [16] Wikipedia, <http://www.wikipedia.org>, accessed 2007-08-22.
- [17] Wikispecies, <http://species.wikimedia.org>, accessed 2007-08-22.
- [18] Wilson, E.O., The Encyclopedia of Life, in: TRENDS in Ecology and Evolution, Vol. 18, No.2 February 2003
- [19] Wilson, E.O., TED Prize wish: Help build the Encyclopedia of Life, <http://www.ted.com/index.php/talks/view/id/83>, accessed 2007-09-03

Author(s):

Anwar Us Saaed, MSc Systems Engineering
Graz University of Technology, Knowledge Management Institute
Inffeldgasse 21a, 8010 Graz
anwar.ussaeed@student.tugraz.at

Alexander Stocker, Mag.
Know-Center
Inffeldgasse 21a, 8010 Graz
astocker@know-center.at

Patrick Hoefler, DI (FH)
Know-Center
Inffeldgasse 21a, 8010 Graz
phoefler@know-center.at

Klaus Tochtermann, Univ.-Prof. Dr.
Graz University of Technology, Knowledge Management Institute; Institute for Networked
Media, Joanneum Research; Know-Center
Inffeldgasse 21a, 8010 Graz
ktochter@know-center.at