Management of Environmental Quality (in press)

ILMAX: A System for Managing Experience Knowledge in a Long-Term Study of Stream Ecosystem Regeneration – an Application of Ecological Informatics

Michael Neumann[1*], Joachim Baumeister[2] & Frank Puppe[2]

[1] Institute of Ecology, Department of Limnology; University of Jena, Carl-Zeiss-Promenade 10, D-07745 Jena, Germany
[2] Institute of Computer Science, Department of Artificial Intelligence and Applied Computer Science; University of Würzburg, Am Hubland, D-97074 Würzburg, Germany

*Author to whom all correspondence should be addressed: Tel: +49-3641-642742; Fax: +49-3641-643325; email: m.neumann@uni-jena.de

KEYWORDS

Restoration ecology; river; online content management system; iZone, knowledge management system;

ABSTRACT

We present the concept and first experiences for a web-based ecological knowledge management system ILMAX. It manages experience knowledge resulting from an ecological long-term investigation of the regeneration process of the stream Ilm (Thuringia, Germany). The tools used are iZone, a web-based content management system with an emphasis on information retrieval and feedback, and D3, a tool kit for building knowledge-based systems and for managing case bases. The difficult initializing phase of knowledge management systems was done by a “knowledge champion” gathering and structuring various kinds of text documents, data sheets and unpublished knowledge of domain experts.

INTRODUCTION

Ecological Informatics provides a wide range of methods to support ecological investigations. While ecosystem analysis, synthesis and forecasting in the past ten years was very much influenced by the introduction of high-performance computing (Recknagel, 2003), nowadays advanced computational technology is used for the elucidation of principles of information processing at and between all levels of complexity of ecosystems (Recknagel, 2003). This links data storage, retrieval and visualization with ecosystem analysis, synthesis and forecasting.

Long-term investigations in ecology collect a large quantity of data, knowledge and experience. When such an investigation starts, it is usually funded only for a few years and it is not evident that the investigation would continue for many years. Therefore, the need to manage experience knowledge is not obvious from the beginning. Most ecological cognition is stored only in informal ways such as text documents, data sheets or even as unpublished knowledge of a domain expert. Furthermore, larger studies are divided into subprojects and scientists usually only work for a few years in such a subproject. All this makes it urgent to develop methods for managing experience knowledge during long-term projects.

Managing experience knowledge can help to structure and retrieve ecological knowledge gathered over a long period by different scientists. The internet offers the capability to store and maintain information from different locations and it provides all the involved scientists with access to the content. The data storage, retrieval and visualization in an information system enables us to integrate data across ecosystem categories, reduce maintenance complexity and make the knowledge available in a such way that it can be used in further analysis. Our long-term aim is to use the structured knowledge to develop a conceptual simulation model of the investigated ecosystem. In this paper we want to present the usage of an online content management system as a tool for data storage, retrieval and visualization.

AVAILABLE EXPERIENCE KNOWLEDGE

Introduction to the Ecological Research Project

Many ecosystem communities are affected by activities of humans or have even been degraded for many years. The types of disturbance range from mechanical influences to chemical contamination of the ecosystem. The appearance of the adapted community differs profoundly from the natural one. When, after many years, the degradation is ameliorated, the community will respond to this change with a succession. This process is not influenced by humans and is therefore called regeneration of the ecosystem. The subject of the ecological investigation project mentioned here is whether the ecosystem can attain its natural state again and how the regeneration can be influenced.

The project “Regeneration- and Functioning-Analysis of Degraded Ecosystems” (Deutsche Forschungsgemeinschaft: GRK 266/1-96) has been investigating since 1996 (until 2005) the regeneration of the stream Ilm (Thuringia, Germany). This is the continuation of an investigation lasting from 1991 to 1994 (BMBF 0339310F). Overall we have seven independent institutes and around 35 scientists involved in the aquatic part of this project.

Study Area

The river Ilm is 130 km long with a catchment area of 1043 km^2 and a mean-flow discharge at the mouth of 5.9 m^3 s^{-1}. The region at the headwater is 75% woodland but the proportion used for agriculture increases along the stream, reaching 95% in the mouth area. The rainfall decreases from the hilly spring region, with 1200 mm a^{-1}, down to 550 mm a^{-1} in the lowland mouth region. In the catchment there are approximately 180,000 residents, living in rather small villages and settlements. After the political reunification of Germany in 1990, many industrial production sites were shut down and modern waste-water treatment plants were built. This improved the water quality rapidly and the regeneration of the aquatic ecosystem has been investigated since then.

Available Experience Knowledge

Experience knowledge was available for the analysis of the contamination, the functionality and the regeneration of the ecosystem. At the beginning of the research the main focus was to document the contamination and the abiotic factors as well as to gain a basic understanding of the ecosystem. Of course, the recovery of the aquatic community was investigated very intensively, because of the fast regeneration process; the retention and turnover characteristics of the stream were investigated in the second phase. In the concluding phase the aim is to integrate knowledge about the separate mechanisms.

For example, during the period of high organic contamination the energy transfer in the ecosystem was investigated. The exponential growth of bacterial populations and therefore an intensification of the decomposition of organic matter can be understood as a defensive reaction against organic pollution (Schoenborn, 1998). Later, the dominance of macrophytes and their response to frequent disturbances of the macrophyte assemblage by floods were studied in the river Ilm (Ensminger et al., 2000). Disturbances by floods also resulted repeatedly in a washout of benthic organisms. The recolonization was investigated by exposure of gravel-filled baskets and was found to be rapid, as an adjustment to the discharge regime of the river (Elser, 1999). These examples show that the available knowledge is divided among separate but not independent investigations.

Problem

In the project considered here, most investigations were done during 21 PhD studies since 1991. Thus, the results and the experience knowledge of this long-term investigation are strongly partitioned. Most of the results have been published in PhD theses or are contained in separate publications. Those attempting integration are confronted with the problem that this knowledge is only informally available; e.g. it is stored as text documents, scientific figures, data sheets or multimedia files. The process of storage, retrieval and visualization of such knowledge is very time-consuming and strongly depends on the authorized scientist. If this particular scientist leaves, the integration process and the expert knowledge are often lost for the project. Therefore, it was necessary to provide a tool that supports the process of storage, retrieval and visualization and ensures that the contents are available to everybody involved in the project.

A second major subdivision is only available as expert knowledge. In ecology it typically takes many years to fully understand the relationships and the functionality of the investigated ecosystem. For this reason, some knowledge only reaches the state of uncertain or soft knowledge. Most scientists will not publish such soft knowledge and when they leave a long-term investigation project this information is lost for the project. On the other hand, during further investigation and by adding more facts, soft knowledge could turn into documented experience knowledge. For this reason, we need to provide an easy-to-use tool that can be applied to collect items of knowledge from scientists, even if they are no longer on-site.

Managing Experience with a Knowledge Management System

In the previous sections we introduced the problem domain and described the existing sources of experience knowledge. We propose to apply a knowledge management system (KMS) for processing experience knowledge. From a technical point of view, the core of the proposed KMS is a content management system (CMS) enriched with formalized knowledge for special purposes (Dieng et al., 1998).

Due to the separation of content, structure and layout, content management systems emphasize reusability and transparency when creating, editing, retrieving and publishing all kinds of content. They have more structure than a list of document files, but less than formal knowledge-based systems and are very useful for organizing heterogeneous pieces of information for a community of practise (CoP) like the degraded stream ecosystem (DSE) community. In the following, we discuss technical solutions for the requirements of the DSE community:

- Organization of heterogeneous documents in a CMS
- Information Retrieval: Browsing and search facilities
- Community aspects and incremental authoring
- Data bases and case-based reasoning
- Inferential formalized knowledge

We also show how these requirements can be achieved by the CMS iZone and the diagnostic shell D3 (Puppe, 1998) developed by the Department of Artificial Intelligence and Applied Computer Science in Würzburg.

Organization of Heterogeneous Documents in a CMS

The central parts of the DSE-CMS are existing documents like dissertations, publications, internal reports, pictures and other multimedia documents, and tables or data bases with empirical data. Since a conversion of all documents to a standard format is very time consuming if not impossible under project conditions, an economic alternative is to enrich the documents with meta-information for retrieval purposes (cf. next section) and to integrate viewers for the original documents. If the documents are very large and contain different kinds of information, they should be split into smaller homogeneous parts. The following kinds of meta-information should be added manually:

- A heterarchical categorization of the domain so that the documents can be classified according to one or in exceptional cases to several categories. The categories will be used for browsing.
- A summary of each document, enabling the user to assess the purpose and content of the original document, since the documents might be rather large. The summary should also contain links to similar or complementary documents.
- If the summary already contains the main keywords for each document, then they can be extracted automatically, otherwise additional keywords should be provided to improve the search.
- A lexicon with important terms and synonyms, as well as an index of authors with a short description are also useful for improving word-based document retrieval.

The CMS stores the text in an XML-structure with tags for title, abstract, full text, author, etc. The text inside the tags is written in HTML (XHTML). The CMS provides tools for automatic generation of word vectors for the documents and for link generation for predefined terms occurring inside the HTML text.

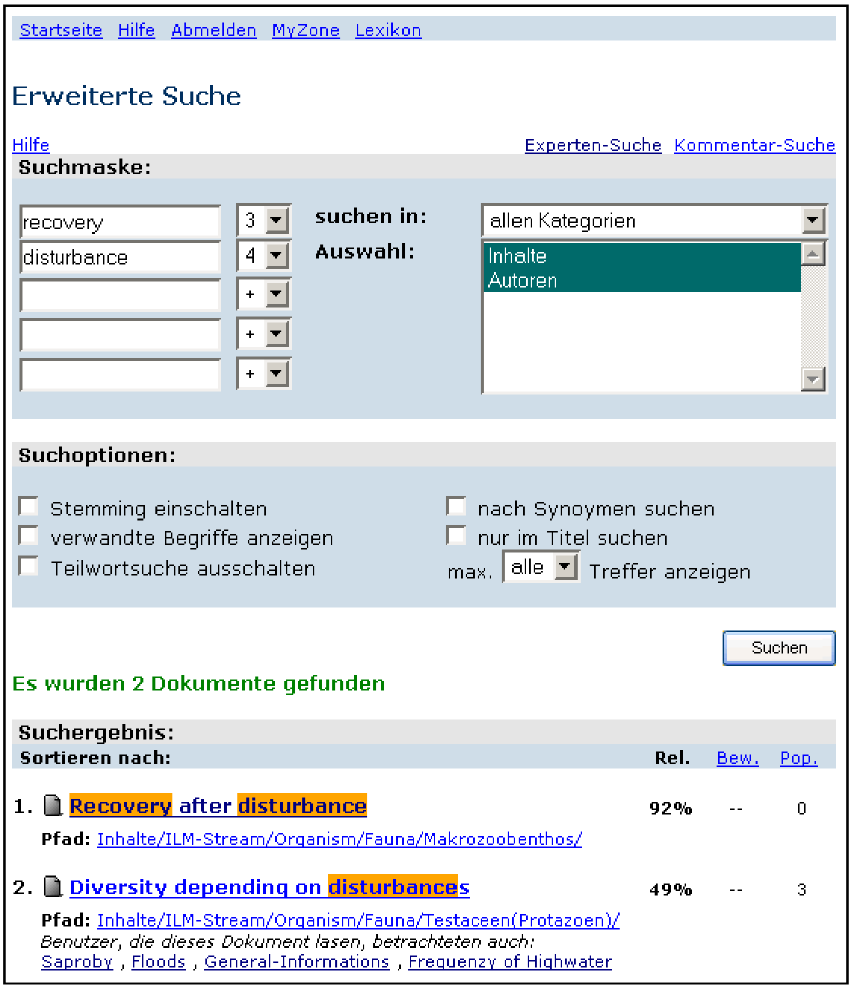

Figure 1: Search in iZone. The user can specify weights (defined by numbers) for the single search terms (e.g. “recovery”, “disturbance”). Furthermore, the system can use stemming and will look for synonyms or related terms. Search results are presented.

Intelligent and Simple Retrieval

One of the most important features of a web-based CMS is the easy and intelligent retrieval of previously stored content to support browsing and searching. Browsing is based on the heterarchical categories defined as meta-information associated with the text documents. The categories are displayed in a permanently visible window as a dynamic hierarchical tree-structure. They are not only a fast navigation mechanism, but also provide a quick overview of the breadth and depth of the contents. Searching is based on the automatic generation of word vectors for the full documents, which is done at compile time for efficiency reasons. Searching requests can be formulated by entering words, Boolean connectors (and, or, not), weightings (how important different words are inside the search request), categories (restricting the search to parts of the heterarchy used for browsing), and wild characters (used to search for similar words simultaneously). The result of the search is a list of documents ranked according to the degree of conformance (computed as a number between 0 and 100). The requested words are highlighted inside the resulting documents.

Community Aspects and Incremental Ways of Authoring

A web-based CMS provides many opportunities for participation of all members of the community of practise (CoP). They include:

- Rating the usefulness of the documents by the authors on a familiar scale, e.g. by school grades.
- Commenting on or annotating the documents, e.g. giving reasons for the school grades, correcting errors or misleading presentations, arguing about controversial opinions, giving examples, adding details, providing links to important material not yet linked or cited, etc.

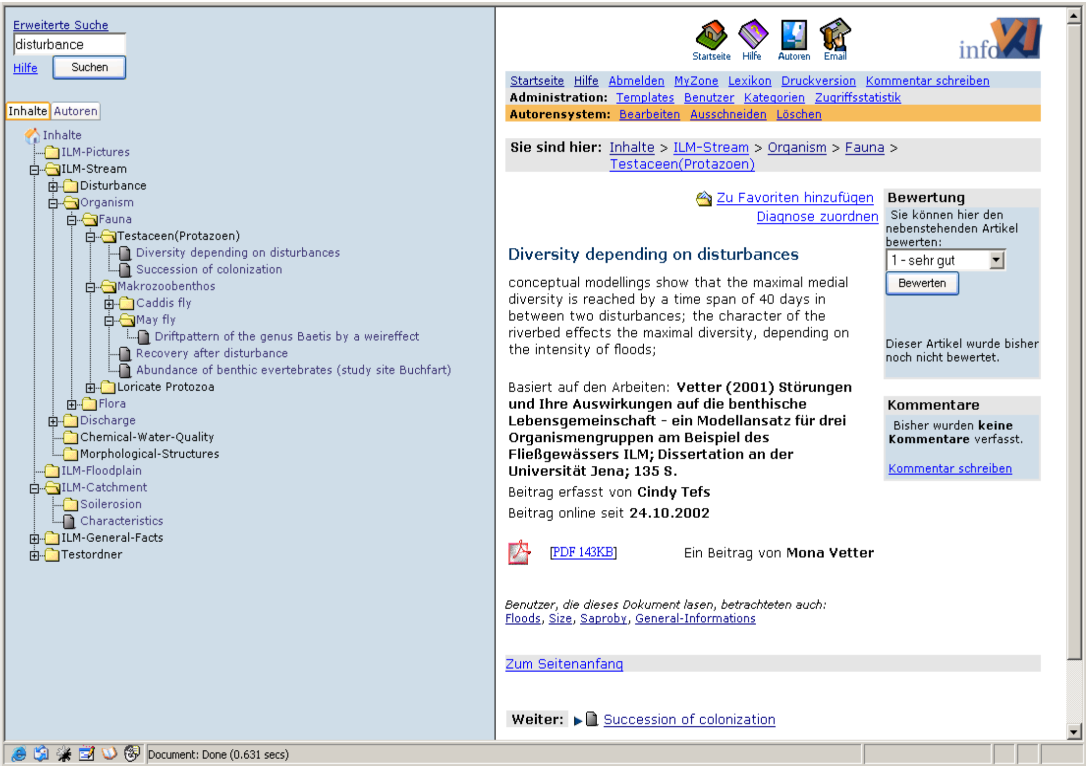

Figure 2: Web-Content Management System iZone. On the left side of the window, categories are arranged in hierarchical order. On the right side, an article about "Diversity depending on disturbances" is presented, summarizing a quite large document in PDF-format.

In a CoP such grades and comments should be personalized, which is implemented by requiring an individual login for all participants entering feedback. The feedback to the articles can be viewed by all other users of the CMS immediately after entering it. However, the typical user would not check all documents for new comments. Therefore the CMS offers a push-facility, where a user can mark documents or categories. Then, whenever a new comment or grade is entered, the user is notified by an email message.

The next step of participation concerns authoring. It should be very easy for authors to write articles, but on the other hand, there should be some organized reviewing process for new articles. For entering new articles we provide both online and offline authoring:

Offline authoring means that the author can write his or her article with a familiar text editor. To enable an automatic conversion to the XML/XHTML format used by the CMS, templates should be offered to the author. The tags of the XML-format (e.g. category, title, abstract, text, author; see above) should be automatically inferred from the document templates. In addition, the text editor should offer an automatic conversion to HTML. Both are implemented e.g. for the Microsoft Word™ text editor. The CMS provides an upload mechanism for the author by file transfer.

Online authoring means that the author edits the document directly without any uploading mechanism. In our CMS this requires some knowledge about XML and HTML. Even without such knowledge, at least small corrections in existing documents can be done quite easily.

A major problem of web-based CMSs is to overcome the difficult initializing phase. We choose the alternative with a knowledge champion (Smith & Farquar, 2000), who provides the initial content, presents the system to the community and is responsible for the quality of the content. However, soon thereafter other authors should participate. Therefore a review process should be implemented. Smith and Farquar mention four alternatives:

- No formal validation: Each author is trusted.
- Validation by personal reputation: Each article must contain an author name.
- Validation by approval by other community members: This can be implemented by the ranking mechanism. An article is accepted if a minimal number of community members have rated it with a predefined score. They are alerted to newly entered articles by email as described above. If the articles should not be freely accessible before the reviewing process is finished, then the articles are made visible only to the reviewing community. The articles are automatically released to the public once enough positive grades have been entered. Since a personalized user management is already necessary for entering the grades, it can be easily extended for this purpose, too.
- Validation by the knowledge champion or his or her agents. This can be best implemented, if all articles being uploaded are not immediately presented in the CMS, but first forwarded to the knowledge champion for approval in a small workflow model.

Since our CMS iZone offers the technical prerequisites for all alternatives, they can be combined or changed during the evolution of the CMS.

Data Bases and Case-Based Reasoning

While most content is available in text documents, there is also more structured data and knowledge, which should also be integrated into the CMS, turning it into a knowledge management system (KMS). An important component is data stored in data bases gathered from field experiments. This is a typical component for an ecological area of application. During the long-term investigation considered here, data were measured repeatedly at specific sites and dates. Of particular interest was the measurement of the contamination of the stream and the aquatic community.

These data have a uniform structure and were gathered recurrently. To avoid perennial and lasting input for the community members we offered a form especially for these types of data effectuate that these data are fully integrated within the KMS. Integration means providing tools for viewing, editing and indexing the data so that this kind of content is fully integrated into the search and browsing routines of the CMS. Besides the convenient data acquisition, another advantage is the structuring of the data during the input process. This is done via the internet directly by the user. Useful inferences on such structured data are plausibility tests of the input data, case-based reasoning (Aamodt & Plaza, 1994) to find similar documents for a new document or statistical and data mining techniques (Han & Kamber, 2000) to extract patterns or rules from the data.

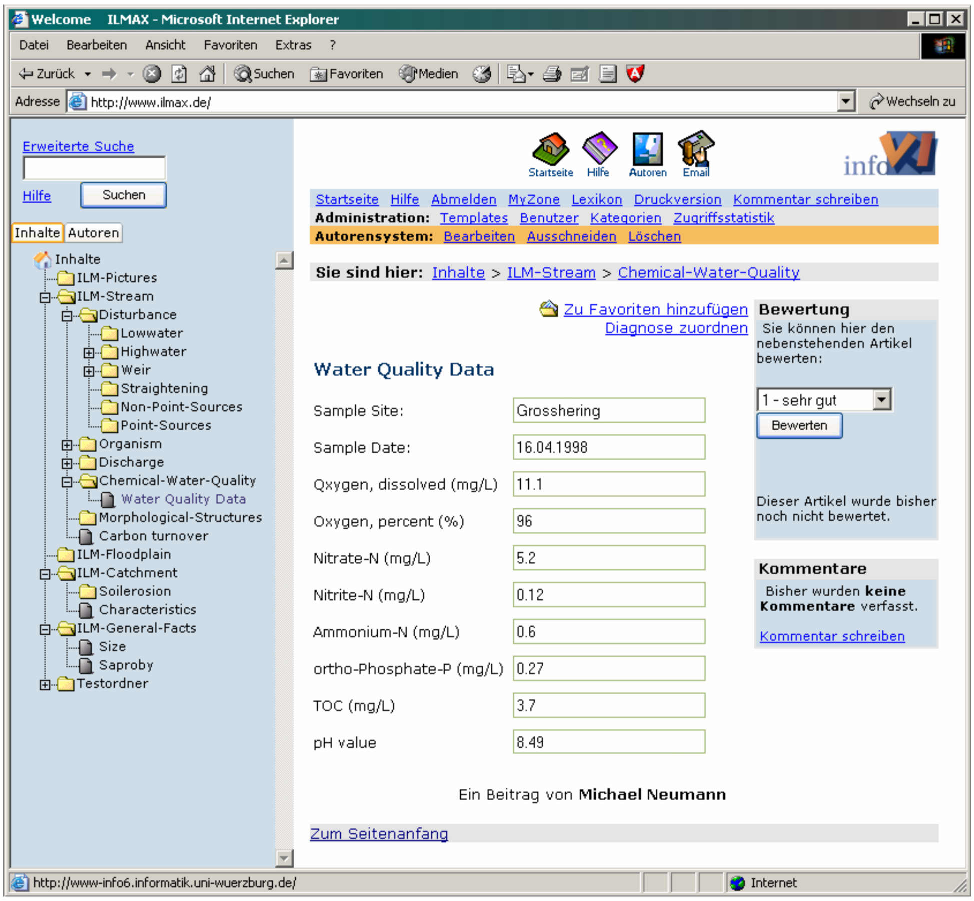

Figure 3: Web-based form for a structured data entry integrated into the Content Management System. The screenshot gives an exemplary data entry of basic water quality measurements.

Inferential formalized knowledge

Knowledge-based systems (KBS) represent another important component of KMS. A major advantage of the CMS iZone is the full integration of KBS. The active linking allows a KBS to run inside the CMS. This means that a KBS is another kind of content of the CMS, which can be viewed, indexed and used. For this, the user can run cases and infer solutions provided by the KBS. In addition the community members are invited to contribute data on their own.

We have built such a KBS in a slightly different domain using the shell-kit D3 (Puppe, 1998), named LIMPACT (Neumann et al., 2003a; Neumann et al., 2003b). This system assesses the overall pesticide contamination of a stream. It interprets biological indicators (taxa of benthic macro invertebrates) to diagnose small streams. When a user runs the system, he has to enter his own abundance data. This would provide an additional case to the CMS.

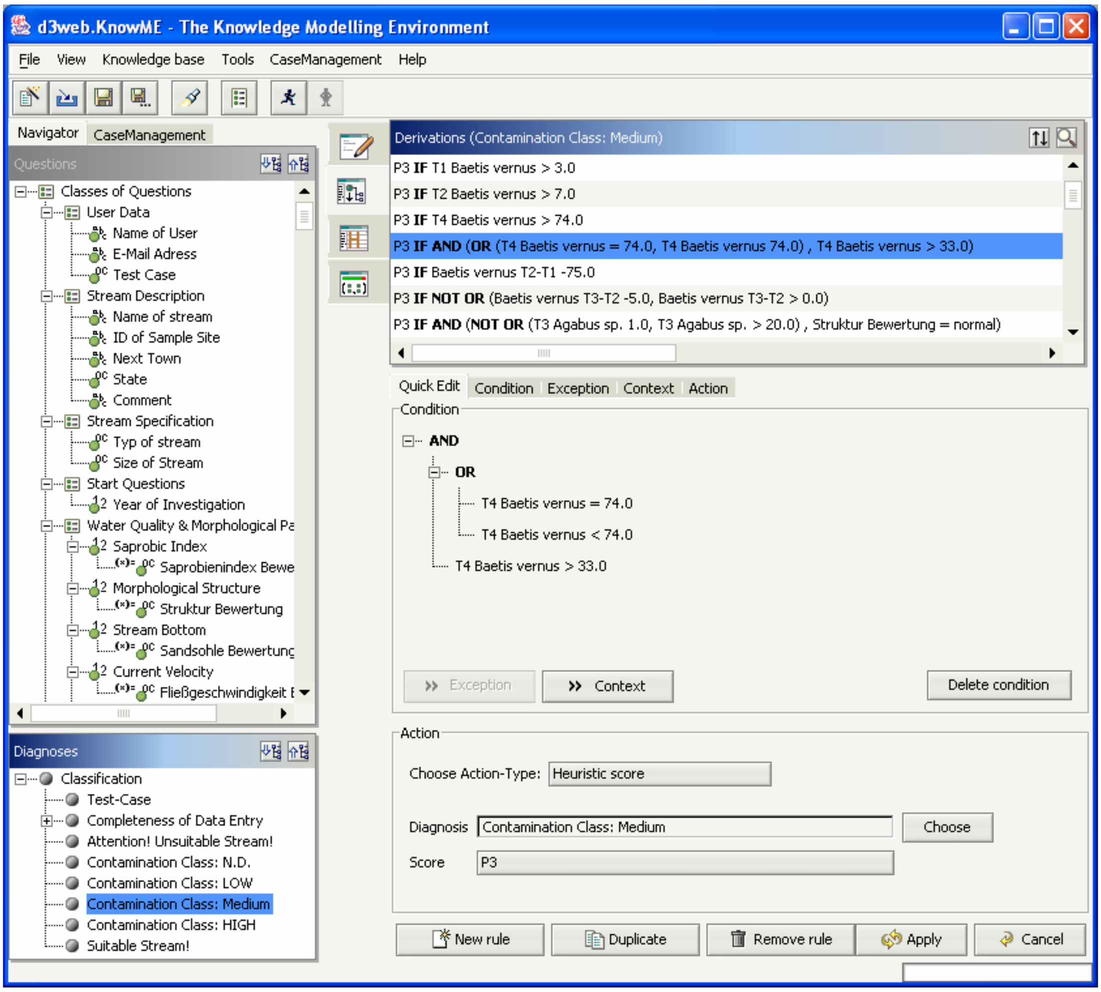

Figure 4: Visual knowledge development tool for creating explicit expert knowledge (rules). The screenshot shows an example of the knowledge base of LIMPACT.

The full integration of KBS into CMS also allows the maintenance and extension of the knowledge bases. The domain expert is able to build knowledge-based systems by formalizing his knowledge via the internet. For this, a knowledge development environment is fully integrated in the CMS.

It is well known that a major bottleneck for knowledge-based systems is the effort expended in formalizing and entering the knowledge. To widen the bottleneck, we use a completely graphical knowledge acquisition tool requiring no programming experience for use. Instead, the domain expert fills in forms, graphs and tables. However, the most difficult part is to formalize knowledge. Therefore, we offer different knowledge representations and knowledge formalization patterns, from which the expert can choose the most suitable (Puppe, 2000). A comparison of heuristic and causal (set covering) knowledge in an ecotoxicological domain can be found in (Neumann & Baumeister, in press).

In the domain considered here, we had many domain experts in the community representing a wide range of non-formalized expert knowledge. A major field of the expert knowledge was the connection between contamination of the stream and the aquatic community of the stream. For this reason, we integrated different KBSs, which are currently under development.

Discussion

Ecological informatics is rather new in the field of ecological science and represents the computational approach to ecosystems analysis, synthesis and forecasting (Recknagel, 2003). It includes the use of information technology to protect the environment or the development of technologies that imitate biological principles, such as evolutionary or genetic algorithms (Keil-Slawik & Brennecke, 1995). It mainly focuses on the higher scales from the organism through the community to the ecosystem scale. The development of knowledge-based systems is a well established area in computer science research. Recently, knowledge management systems have been formed as a synthesis between the knowledge-based and information system communities.

However, in the field of ecosystem research, according to our information, only very few approaches to knowledge management systems have been published. In other domains like corporate investigations or medical informatics, knowledge management systems are well established methods. This encourages us to believe that knowledge management systems can also be applied successfully in ecological science.

The use of an online content management system in the aquatic ecosystem research is not documented, at all. We named our system ILMAX (from river ILM and Managing Experience Knowledge) and it is available at http://www.ilmax.de via the internet. Guests may read, but are not allowed to publish or edit content. Instead they have an opportunity to rate the usefulness or to comment on the content. Although ILMAX is still a prototype, we can already report some preliminary experiences.

Our decision to start the initializing phase with a knowledge champion proved to be successful. He released content out of publications and started writing the lexicon. In a second step he informed all members of the community and explained the idea of ILMAX in detail. He convinced the members to invest some of their work time to publish their results, data and knowledge. The closer in space and time the scientists were still related to the project, the easier this motivation process was. However, the community was very heterogeneous. One former scientist, who had already left the project, even published some content from South America. Only one older scientist had only very little experience with the internet and needed more explanation and help. The quality of an online content management system, in general, is strongly correlated with the depth of the content. This can only be provided by true experts writing as authors of the system. Rating and comments by guests may complete the knowledge acquisition, but only provide flat content.

A general problem for scientists is the publishing of not yet published results or of uncertain (soft) knowledge. In the first case we need to limit access to the information system to internal members of the investigation project only. This is no technical problem. For the second case it was difficult to convince the scientist that small pieces of knowledge are also very valuable to the complete system and are worth publishing. Overall, we found a good willingness to support the development of ILMAX, which can be explained by the large size of the ecological project and the large number of scientists involved.

We now hope that the scientists will frequently use ILMAX as an information system for their own purposes. The authors of this paper want to convert the content of the systems into a conceptual model for the regeneration process of degraded stream ecosystems.

CONCLUSIONS

- In long-term ecological studies the use of an online content management system can help to store, retrieve and visualize experience knowledge.

- The internet facilitates collection of items of data and expert knowledge from scientists, even if they are no longer on-site.

- The process of knowledge acquisition in an online content management system generates an information system that can be used by all involved scientists and no longer depends on the knowledge champion.

- The structuring of the knowledge enables an advancement, for example, to develop a conceptual model.

ACKNOWLEDGEMENTS

This study was supported by grants of the Deutsche Forschungsgemeinschaft DFG within the graduate research group “Functioning- and Regeneration Analysis of Degraded Ecosystems” (GRK 266/1-96).

REFERENCES

Aamodt, A. & Plaza, E. (1994), Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches, Artificial Intelligence Communications, 7 (1), 39-59.

Elser, P. (1999), Use of colonization baskets for the investigation of disturbance phenomena in streams under model conditions, Limnologica, 29, 120-127.

Ensminger, I., Hagen, C., & Braune, W. (2000), Strategies providing success in a variable habitat: I. Relationships of environmental factors and dominance of Cladophora glomerata, Plant Cell and Environment, 23, 1119-1128.

Dieng, R., Corby, O., Giboin, A. & Ribiere, M. (1998), Methods and Tools for Corporate Knowledge Management, in Proceedings of KAW'98: The Eleventh Workshop on Knowledge Acquisition, Modelling and Management from 18th to 23rd April, Banff (Canada).

Han, J. & Kamber, M. (2000), Data Mining: Concepts and Techniques, Morgan Kaufmann, San Mateo (USA), 550 p.

Keil-Slawik, R. & Brennecke, A. (1995), Ökologische Informatik - Alternatives Leitbild oder unerfüllbares Wunschbild? FIfF-Kommunikation, 12, 18-22.

Neumann, M. & Baumeister, J. (2004), A rule-based vs. a model-based implementation of the knowledge system LIMPACT and its significance for maintenance and discovery of ecological knowledge, in Lek, S., Scardi, M., Verdonschot, P. Jorgensen, S.E. & Park, Y.S. (Ed) Modelling community structure in freshwater ecosystems, (in) Springer-Verlag, Berlin.

Neumann, M., Baumeister, J., Liess, M., & Schulz, R. (2003a), An expert system to estimate the pesticide contamination of small streams using benthic macroinvertebrate as bioindicators, Part 2: The knowledge base of LIMPACT, Ecological Indicators: 2 (4) 391-401.

Neumann, M., Liess, M., & Schulz, R. (2003b), An expert system to estimate the pesticide contamination of small streams using benthic macroinvertebrate as bioindicators, Part 1: The database of LIMPACT, Ecological Indicators: 2 (4) 379-389.

Puppe, F. (1998), Knowledge Reuse among Diagnostic Problem Solving Methods in the Shell-Kit D3, in International Journal of Human-Computer Studies 49, 627-649.

Puppe F. (2000), Knowledge Formalization Patterns, in Proceedings of PKAW 2000: The Pacific Rim Knowledge Acquisition Workshop from 11th to 13th December, Sydney (Australia).

Recknagel, F. (Ed.) (2003), Ecological informatics: understanding ecology by biologically inspired computation, Springer-Verlag, Berlin, 398 p.

Schoenborn, W. (1998), Changes of biological time patterns and of the energy transfer on exposed slides and stone surfaces (boundary layers with fractal structures) by organic pollution in an oligotrophic stream (Ilm, Thuringia, Germany), Limnologica, 28, 347-361.

Smith, R. & Farquar, A. (2000), The Road Ahead for Knowledge Management: An AI Perspective, AI-Magazine, 21, 17-40.