

Visualizing Research Works in the Water Resources Industry

Ricardo Kawase, Patrick Siehdnel, Ujwal Gadiraju

L3S Research Center, University of Hannover
30167 Hannover, Germany
{kawase, siehdnel, gadiraju}@l3s.de

Abstract—With the increasing practice of making data openly available, nowadays there is a growing amount of information easily available pertaining to water resources and ecology. Scientific works by researchers across the world contribute to the abundance in such data. Major challenges that emerge due to the volume of data include the discovery of useful and relevant content, as well as learning and interpretation of the various disparate content. In this work, we aim to aid researchers and interested stakeholders in understanding the vast landscape of scientific research in the water resources industry. We integrate different sources of data from the Web; journals from Elsevier, tweets from Twitter, and Wikipedia annotations. We use interactive visualizations in order to engage the users and satisfy their information needs.

Targeted Content Track: Water Resources and Ecology

I. INTRODUCTION

As described through the aims of the Water Resources and Industry journal¹, there is a gap in the role that the water resources industry plays in the exploitation, management and treatment of water resources. Various industries participate to different extents in treating or disposing the waste water, resulting in the corresponding manufacturing processes. Depending on the geographical location of the industries, there is a varying impact on the local resources. Therefore, different approaches are required in order to cater for the water resources management.

Several researchers around the world investigate approaches in order to realize optimal solutions for the most efficient methods to manage water resources. An important aspect in this regard, is that breakthroughs in a particular field are not restricted based on geographical boundaries. This goes hand-in-hand with the principles of openly accessible data, so that researchers can learn from one another and thereby hasten the process of developing suitable methods for the management of water resources. Some of the major challenges that emerge due to the volume of data that is hence shared across different related domains, include the discovery of useful and relevant content, as well as learning and interpretation of the various disparate content. In this work we propose a system that helps researchers to discover and understand the vast landscape of scientific works in the water resources industry. In order to

achieve this, we integrate different sources of data from the Web; journals from Elsevier, tweets from Twitter, and annotations using a toolkit called WikipediaMiner², for tapping the rich semantics encoded within Wikipedia. By combining data from these disparate sources, we enrich the a user's experience of satisfying her information need. The content and context of journals provide sounds scientific pointers, while annotated entities enhance a the user's understanding of the subject. Finally, relevant tweets with regards to the search query of a user enable the user to observe the present day contexts of the query subject at hand.

The remainder of this document is structured as follows. In the next Section, we describe our system and the underlying technologies used to build it. Section III portrays the key characteristics of our system, while highlighting the distinguishing features with respect to existing standalone systems. We finally draw conclusions in Section IV, and provide additional screenshots of the interface in the Appendix, Section V thereafter.

II. SYSTEM DESCRIPTION

Our system can be accessed at the following url; <http://www.l3s.de/~kawase/vici/>.

A. Interface and the underlying Data

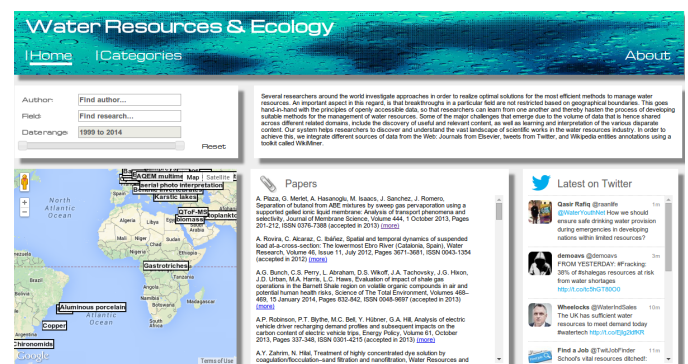


Fig. 1. Screenshot of our interface showing the home screen.

Our system relies on a MySQL backend and the interface has been developed using PHP, HTML and javascript. As a

¹<http://journals.elsevier.com/water-resources-and-industry/>

²<http://wikipedia-miner.cms.waikato.ac.nz/>

first step we utilized the data from relevant Elsevier journals from the year 1999 till 2013. We extracted relevant metadata such as author names, abstracts, full text, and keywords from the focused track corpus from Elsevier.

B. Research Around the World

In order to provide the users with an understanding and an overview of where each research is being done, we use the Google Maps API³ integration. From the metadata that contains the address of each author, we query the API to retrieve the corresponding coordinate. On these coordinates, we place the on the map the related fields of research (author given keywords), as shown in the Figure 2.



Fig. 2. A map with keywords pertaining to research works produced in the corresponding geographic area.

C. Dynamic Tweet Timelines

In order to provide users with a real-time view of related events to given topics, we use the Twitter API to support our system with dynamic timelines, that can help a user to analyze the present context of the keywords within her query. This combination of data from scientific works and data from a social network help to enhance the user’s overall experience in satisfying her information need. This is achieved by leveraging both the elementary and exploratory features of the interface. Note that Twitter does not provide dynamic search widgets. In order to provide users with such functionality, we implemented a backend PHP connection to Twitter’s search API that enables us to dynamically query for different results. The results are then presented to the users in a familiar way, utilizing HTML and CSS templates.

³<https://developers.google.com/maps/>

D. Categories

We use the Wikipedia category graph in order to enrich the knowledge from scientific papers. We specifically consider the top-level categories to be (i) Food, (ii) Water, and (iii) Countries (by continent) as suggested by the stakeholders of the datasets. First, we annotate all the abstracts and the highlights of papers with corresponding Wikipedia pages using WikipediaMiner. Finally, we relate each scientific paper based on the annotations of the abstract with categories in the Wikipedia category graph, following methods introduced in previous works [1]. These categories can include the 3 top-level categories mentioned already and all possible sub-categories through varying depths of the Wikipedia category graph.

1) *Determining appropriate Categories:* For the annotation we separated the abstracts on a sentence level and used the remainder of the abstract as context. Providing meaningful context increases the quality of annotations for each sentence. Finally, we relate each scientific paper based on the annotations of the abstract with categories in the Wikipedia category graph. The related categories for each article are generated using a top-down approach on the Wikipedia category Graph. Starting from the 3 selected top categories we follow each child category down to a distance of 3 hops in a recursive fashion. Since the Wikipedia category graph does not have a simple tree structure, but can also contain cycles, we only followed categories that have a larger or equal distance to the root category. A more detailed description of the proposed method is given in [1], [2]. The articles belonging to the visited categories are then related to all categories that were visited in the same branch of the graph.

TABLE I
NUMBER OF CATEGORIES.

Category Description	Total Number
Top Categories	3
Second-Level Categories	88
Third-Level Categories	1027

In total there were 5738 annotations within the abstracts of the papers. This annotation process comprised of links to 4964 distinct Wikipedia pages. With the hierarchical categorization, we finally establish 5112 relations from the scientific papers to the categories.

In the end, categories provide a structured browsing interface that supports users in finding topic-relevant papers.

E. Snippets as Category Pointers

While browsing papers through categories, a user may ask herself why a given paper belongs to a given category. To fill this gap, we extracted snippets of the papers with respect to each annotation found. Thus, users can retrieve the snippets of papers due to which the papers are categorized accordingly, by clicking on ‘*Why this category?*’. This greatly benefits a user since one can easily assess whether a paper is of interest or not within a category by merely skimming through the snippets.

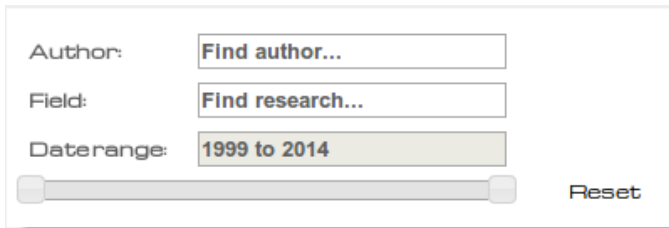


Fig. 3. Search fields embedded within the interface.

III. DISTINGUISHING FEATURES

In this section, we present the key characteristics that we provide to aid users in the exploration of content, as well as discovery of information related to water resources and ecology through our system.

[i]. The interface consists of a search field as shown in Figure 3, where a user can enter a combination of the name of an author and/or the field of research she is interested in. In addition to this, the user can also select the temporal range within which the results are to be retrieved. To aid the user and increase the usability of the system, we support the search queries with *auto-completion*.

[ii]. Scientific papers related to the entered query (either an author, a field, or a combination of both) and within the selected date range, are retrieved and presented to the user in the center of the screen as shown in Figure 4. The complete citation of the paper including the year of acceptance and publication are presented to the user. A user can access the the corresponding online version of a scientific work by clicking on *'more'*.

Papers

José Lailson-Brito, Paulo Renato Dorneles, Cláudio Eduardo Azevedo-Silva, Alexandre de Freitas Azevedo, Lara Gama Vidal, Juliana Marigo, Carolina Bertozzi, Regina Célia Zanelatto, Tatiana Lemos Bisi, Olaf Malm, João Paulo Machado Torres, Organochlorine concentrations in franciscana dolphins, Pontoporia blainvillei, from Brazilian waters, Chemosphere, Volume 84, Issue 7, August 2011, Pages 882-887, ISSN 0045-6535 (accepted in 2011) [\(more\)](#)

Reinhard Heerkloss, Werner Schnese, A long-term series of zooplankton monitoring of a shallow coastal water of the southern Baltic, Limnologica - Ecology and Management of Inland Waters, Volume 29, Issue 3, July 1999, Pages 317-321, ISSN 0075-9511, (accepted in 2005) [\(more\)](#)

Fig. 4. Scientific papers related to the query, in this example, *'coastal waters'*.

[iii]. In addition, we include a Twitter widget on the right-side of the home screen, where the most recent tweets pertaining to the entered query are presented to the user. The user can consume this information and also *reply*, *retweet*, or *favorite* any tweets. We use the Twitter search API to facilitate dynamic search, distinguishing our implementation from typical Twitter widgets where dynamic searches are not feasible.

[iv]. The categories as shown in Figure 6, provide a user with additional metadata about the papers with respect to the context of the scientific works. The user can navigate within subcategories and parent categories. They aid the user with

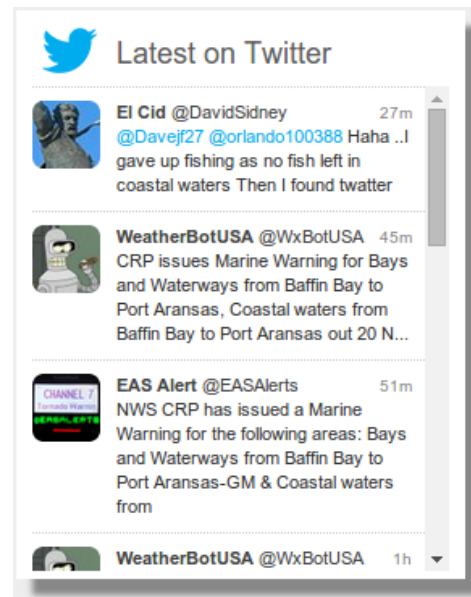


Fig. 5. Twitter widget showing the latest tweets related to an entered field, in this example, *'coastal waters'*.

exploration of various topics as well as with the easy retrieval of information.

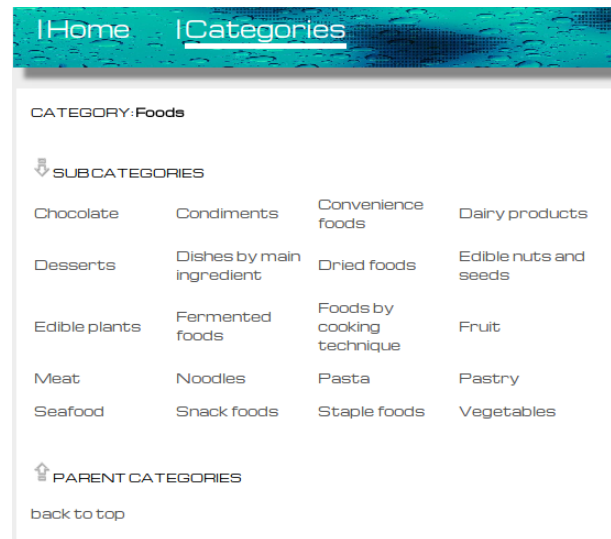


Fig. 6. The top-level category of *Foods* as presented in the interface.

[v]. By using snippets from papers in order to explain to a user why a scientific paper is found within a particular category, we provide a user with an easy and quick understanding of how that paper can satisfy her information need or serendipitous discovery. An example within the category of *'reservoir'* is presented in the Figure 7. By providing text from papers corresponding to the category selected, users can have a preview and understanding of why a paper is categorized as such.

[vi]. By using maps along with keywords describing scientific research works, we make it easy for a user to discover and

Papers (11)

367 Gebre-Mariam Zinabu, The effects of wet and dry seasons on concentrations of solutes and phytoplankton biomass in seven Ethiopian rift-valley lakes, *Limnologia - Ecology and Management of Inland Waters*, Volume 32, Issue 2, July 2002, Pages 169-179, ISSN 0075-9511, (accepted in 2001) ([why this category?](#)) ([more](#))

Paper snippet
...The chemical characteristics and phytoplankton biomass (measured as chlorophyll-a concentrations) of seven lakes and one reservoir in the Ethiopian rift-valley were studied during the wet and dry seasons between 1990 and 2000. ...

Fig. 7. Snippets from the paper due to which a paper is categorized accordingly, in this example, *reservoir*.

connect geographic locations to areas of research. Studying the correlation between geographic terrains, and other landscape-intrinsic features with the popular realms of scientific works produced can be interesting, but is out of the scope of our work. We however, make this avenue easily accessible for further investigation through our interactive graphical interface.

A. Audience

We perceive the largest group of our target audience to be researchers in the broad fields of water resources and ecology, both in industry and academia. Researchers around the globe can easily assess the geographic distribution of specialized realms of research. This can particularly help new and upcoming researchers to collaborate with experts, or shorten their learning curve. Apart from this main target audience, we believe the system can be useful to study the landscape of research in this field by any interested stakeholders, for example, students of these sciences, journalists, and so forth. This system can however, be easily extended in order to include other research communities and cater to a larger audience.

B. Usability

Our system and interface have been designed to comply with general principles of usability. The system provides easy access and intuitive navigation to all sections of the interface, along with supportive help texts and snippets. Additionally, it was developed to be platform independent and to support any modern browser in any device. We also facilitate the access of our system via mobile devices.

C. Innovation in Education

As well known in the field of learning and education, a significant challenge to a learning process is the acquisition of appropriate information that satisfies a learners needs. This is even more apparent when a learner has a different specialization than the field he/she is seeking. While it is not hard to find relevant research using existing tools, there is little one can find beyond explicitly specified authors, corresponding publications, and venues. This has been previously established in the technology enhanced learning community [3].

Consider the context where a researcher who is experienced in a specific field such as *'hydrology'*, is interested to find reliable researchers in the field of *'fluvial landforms'*. Using our system, not only can such an information need of a user be satisfied, but with appropriate links to existing works in

the specific field, visualized through an interactive geographical map, as well as live contexts through tweets, the user-experience is vastly improved.

The main contribution of our work is the assistance that the interface provides to users, for discovering information, relevant related works in fields of interest, and the geographical correlation of specific fields of study. By exploiting the *'Categories'* section of the interface, users can easily find relevant information. The interface further helps to track evolution in the research fields, by leveraging yearly based filters to help users in identifying the transition and development of knowledge and authors along the years.

IV. LEGAL ASPECTS

In our work, we use data that is freely available on the web and data that has been made available for the LinkedUp VICI challenge, i.e., the Water Resources and Ecology corpus⁴. The ToC of our work can be found at the following link <http://www.l3s.de/~kawase/vici/about.php>.

A. Web Interface Terms and Conditions of Use

1) *Terms*: By accessing this system, a user agrees to be bound by these Terms and Conditions of Use, all applicable laws and regulations, and agree that the user is responsible for compliance with any applicable local laws. If a user does not agree with any of these terms, she is prohibited from using or accessing this site. The materials contained in this web site are protected by applicable copyright and trade mark law.

2) *Use License*: Permission is granted to temporarily download one copy of the materials (information or software) on this system for personal, non-commercial transitory viewing only. This is the grant of a license, not a transfer of title, and under this license a user may not: modify or copy the materials; use the materials for any commercial purpose, or for any public display (commercial or non-commercial); attempt to decompile or reverse engineer any software contained on this web site; remove any copyright or other proprietary notations from the materials; or transfer the materials to another person or "mirror" the materials on any other server. This license shall automatically terminate if a user violates any of these restrictions and may be terminated at any time. Upon terminating a user's viewing of these materials or upon the termination of this license, a user must destroy any downloaded materials in her possession whether in electronic or printed format.

3) *Disclaimer*: The materials provided via our system are provided "as is". The interface makes no warranties, expressed or implied, and hereby disclaims and negates all other warranties, including without limitation, implied warranties or conditions of merchantability, fitness for a particular purpose, or non-infringement of intellectual property or other violation of rights. Further, this system does not warrant or make any representations concerning the accuracy, likely results, or reliability of the use of the materials on its Internet web site

⁴<http://linkedup.eu/catalogue/data/static/tct/wrec.rdf.xml>

or otherwise relating to such materials or on any sites linked to this system.

4) *Limitations*: In no event shall this system or its suppliers be liable for any damages (including, without limitation, damages for loss of data or profit, or due to business interruption,) arising out of the use or inability to use the materials on this system, even if the system itself or a web site authorized representative has been notified orally or in writing of the possibility of such damage. Because some jurisdictions do not allow limitations on implied warranties, or limitations of liability for consequential or incidental damages, these limitations may not apply to a user.

5) *Revisions and Errata*: The materials appearing on this interface could include technical, typographical, or photographic errors. The system does not warrant that any of the materials on its interface are accurate, complete, or current. This system may make changes to the materials contained on its interface at any time without notice. This web site does not, however, make any commitment to update the materials.

6) *Links*: This system has not reviewed all of the sites linked to its Internet interface and is not responsible for the contents of any such linked site. The inclusion of any link does not imply endorsement by our system of the site. Use of any such linked web site is at the user's own risk.

7) *Site Terms of Use Modifications*: This system may revise these terms of use for its web site at any time without notice. By using this interface, a user agrees to be bound by the then current version of these Terms and Conditions of Use.

8) *Governing Law*: Any claim relating to this system shall be governed by the laws of the Federal Republic of Germany without regard to its conflict of law provisions.

General Terms and Conditions applicable to Use of a Web Site.

9) *Privacy Policy*: A user's privacy is very important to us. Accordingly, we have developed this Policy in order for the user to understand how we collect, use, communicate and disclose and make use of personal information. The following outlines our privacy policy.

Before or at the time of collecting personal information, we will identify the purposes for which information is being collected. We will collect and use of personal information solely with the objective of fulfilling those purposes specified by us and for other compatible purposes, unless we obtain the consent of the individual concerned or as required by law. We will only retain personal information as long as necessary for the fulfilment of those purposes. We will collect personal information by lawful and fair means and, where appropriate, with the knowledge or consent of the individual concerned. Personal data should be relevant to the purposes for which it is to be used, and, to the extent necessary for those purposes, should be accurate, complete, and up-to-date. We will protect personal information by reasonable security safeguards against loss or theft, as well as unauthorized access, disclosure, copying, use or modification. We will make readily available to customers information about our policies and practices relating to the management of personal information.

We are committed to conducting our business in accordance with these principles in order to ensure that the confidentiality of personal information is protected and maintained.

V. CONCLUSIONS

In this document, we present our system, which is a graphical user interface to explore and investigate the vast landscape of scientific research works in the field of Water Resources and Ecology.

We facilitate the search and retrieval of relevant research works within the Elsevier journals from the year 1999 till 2013. By providing different parameters that users can enter as a part of their query, such as *author names*, *field of research*, and the *date range* of interest, we satisfy the user's information need in an elegant manner by leveraging appropriate filters.

We combine data from the scientific journals alongside tweets from Twitter and present this information to the user in a coherent interface to create a multi-dimensionality of the user's information need. This is more so appreciable in an exploratory search scenario, and can improve the overall user-experience.

Finally, by leveraging the Wikipedia category graph and annotating the scientific papers, we present the user with various levels of categories through which she can explore existing research works, or find interesting works with considerable ease. By providing snippets from the papers based on which papers are categorized accordingly, users can easily skim through various categories until their information need is met with or relevant knowledge is discovered.

The Appendix shows a few selected screenshots of the online demo. We invite the reader to access our system online demo and kindly ask for feedback.

<http://www.l3s.de/~kawase/vici/>

VI. APPENDIX: SCREENSHOTS

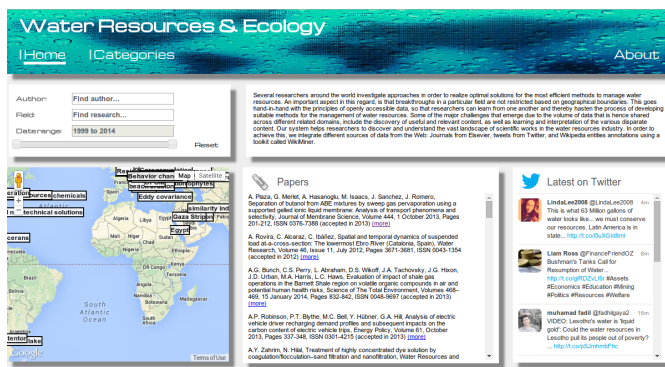


Fig. 8. Home-screen of the Web interface.

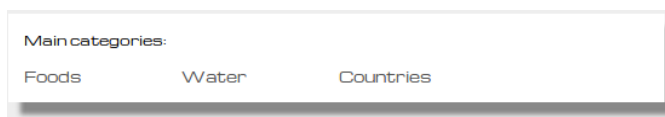


Fig. 9. The main categories presented to a user.

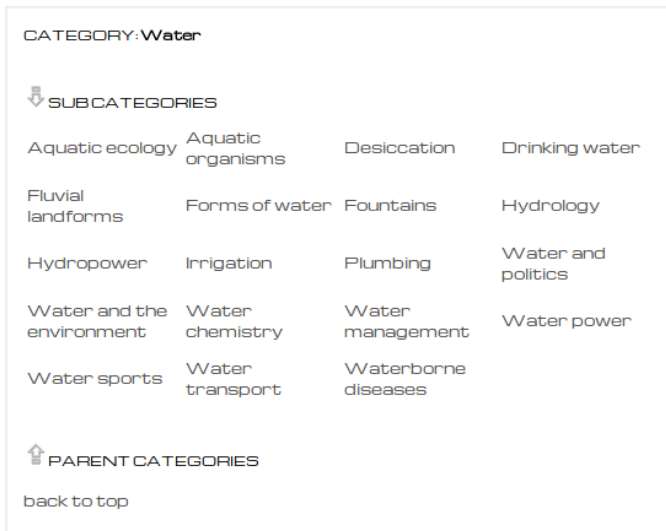


Fig. 10. Example 1 : Level 2 sub-categories of Water.

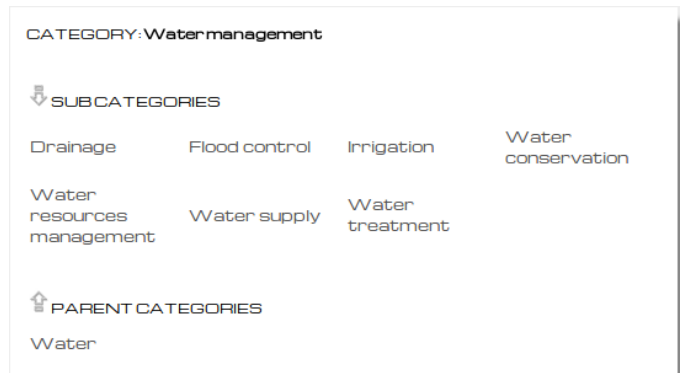


Fig. 11. Example 2 : Level 3 sub-categories of Water Management.

REFERENCES

- [1] P. Siehndel and R. Kawase, "Twikime! - user profiles that make sense," in *International Semantic Web Conference (Posters & Demos)*, 2012.
- [2] R. Kawase, P. Siehndel, B. Pereira Nunes, E. Herder, and W. Nejdl, "Exploiting the wisdom of the crowds for characterizing and connecting heterogeneous resources," in *Proceedings of the 25th ACM conference on Hypertext and social media*. ACM, 2014, pp. 56–65.
- [3] R. Kawase, P. Siehndel, and U. Gadiraju, "Technology enhancing learning: Past, present and future," in *Open Learning and Teaching in Educational Communities*. Springer, 2014, pp. 193–206.