

# Intelligent Support to the Retrieval of Information about Hydric Resources<sup>\*</sup>

Cristina Gena and Liliana Ardissono

Dipartimento di Informatica, Università di Torino  
Corso Svizzera 185, Torino, Italy  
{cgena, liliana}@di.unito.it

**Abstract.** This paper presents the adaptive search features offered by ACQUA, a Web-based system presenting information about the hydric resources of the Piedmont Italian Region. ACQUA enables the user to retrieve qualitative and quantitative data in different formats, supporting direct data manipulation. Moreover, the system supports the user in the search for information by complementing her explicit search queries with follow-up queries frequently occurring together in navigation paths. In this way, the user may retrieve complete information in an efficient way. The paper describes the results of an evaluation of the adaptivity features carried out with real users.

**Copyright by Springer Verlag. This paper is going to appear in the proceedings of the AH 2004 Conference, Eindhoven, 2004, published by Springer Verlag in the Lecture Notes for Computer Science collection (see <http://www.springer.de/comp/lncs/index.html>).**

## 1 Introduction

Public Web sites play a central role in the dissemination of information at low cost, but they frequently fail to present the information in a satisfactory way. For example, most data is presented in static Web pages hardly satisfying individual information needs; see [7]. Moreover, qualitative and quantitative data is typically presented in a format that cannot be directly exploited for processing purposes. Therefore, users cannot directly exploit the retrieved data for interpretation (e.g., to generate histograms and charts, or to compute aggregated data, such as maximum, minimum and mean values).

Indeed, advanced user modeling and content generation techniques have the potential to enhance the interaction capabilities of these Web sites. However, these techniques cannot be easily adopted at the current stage, because of serious scalability and maintenance requirements. However, we believe that some minimalistic Adaptive Hypermedia techniques could be applied to enhance the interaction with the user without overloading the Web sites too much. For example, lightweight techniques could be employed to

---

<sup>\*</sup> This work was funded by Regione Piemonte, Direzione Risorse Idriche. We thank G. Negro, G. Amadore, S. Grisello, A. Giannetta, M. Governa, E. Quinto, M. Demeo and V. Pellegrino, who assisted us during the system development and provided the domain specific knowledge.

REGIONE PIEMONTE **IL MONITORAGGIO DELLE ACQUE**

CORSI D'ACQUA (dati qualitativi) | CORSI D'ACQUA (dati quali-quantitativi misurati in continuo) | ACQUE SOTTERRANEE | LAGHI

ricerca semplice - **ricerca avanzata**

Sei in: Corsi d'acqua (dati qualitativi) - Ricerca avanzata > Caratteristiche della stazione di monitoraggio

[Torna alla ricerca avanzata](#)

**Caratteristiche della stazione di monitoraggio TORINO**

► [visualizza la CTR \(file jpg, 173 Kb\)](#) oppure [scaricala sul tuo computer \(file zippato, 166 Kb\)](#)

**Fiume PO**

**IDENTIFICAZIONE DEL PUNTO**

Codice **001095**  
 Comune **TORINO** Località **PARCO MICHELOTTI**  
 Quota s.l.m. **220 m** Ordine asta **I**  
 Carta Tecnica Regionale Sezione n° **156090 - TORINO NORD EST**  
 Anno di istituzione del punto **2000**  
 Punto di monitoraggio **BIOLOGICO/CHIMICO**  
 Centralina automatica Tipo

| CRITERI DI LOCALIZZAZIONE |  |
|---------------------------|--|
| <b>1. Idrologici</b>      |  |
|                           |  |
| <b>2. Ambientali</b>      |  |

Fig. 1. Portion of the page describing the Torino observation point on Po river.

present data in different formats, so that it can be visualized and processed in heterogeneous software environments. Moreover, simple personalization rules could be applied to speed up the user's search activity, at least in the most common types of search.

In this paper, we describe the content generation and personalization techniques applied in the ACQUA prototype information system, which presents data about hydric resources of the Piedmont Region. The system provides the information requested by the user in pictorial and machine processable formats. Moreover, the system complements the search results with related data usually searched for, in order to speed up the search task. In order to prove the usefulness of the adaptive features offered by our system, we evaluated them with real users. The rest of this paper is organized as follows: Section 2 presents the project within which the ACQUA prototype was developed. Section 3 describes the information about hydric resources managed by the system. The search features offered by the system are described in Section 4 and evaluated in Section 5. Section 6 compares our work to the related one and concludes the paper.

## 2 The Project

In 2003, the Hydric Resources Division of the Piedmont Region and the University of Torino started a project for the development of ACQUA, an interactive Web site presenting information about hydric resources derived from the monitoring activities on the territory. The goal was to present up to date information thus reducing the need to deliver data on a one-to-one basis via e-mail messages and paper publications. The technicians of the Division guided us in the development of the Web site by providing

```

RIVERS: Qualitative data:
        Environmental State parameters:
          Environmental State (SECA): sufficient;
          Ecological State (SECA): class 3;
          IBE index: 7;
        ...
        Biological parameters:
          Acarus (A): 6;
          Biological quality: class 6;
        Chemical, physical and micro-biological parameters:
          Eschericha coli: 8300 UFC/100 ml;
          Total Azote: 9.63 mg/l;
        ...
        Quantitative data:
          Hydrometric level: 0.023 m;
          Average daily capacity: 47 mc/s; ...

```

**Fig. 2.** Sample qualitative and quantitative parameters per resource type.

application requirements and a repository of e-mail messages which they exchanged with users along the years. The messages provided us with evidence about the types of users interested in hydric resources, the inspected information and the regularities in the data search. Most questions came from other Public Administrations, technicians (e.g., companies building bridges, houses, etc.), attorneys (typically interested in data about specific regions, e.g., as a consequence of environmental disasters), farmers (interested in the bio-chemical state of their fields), students (university and PhD).

Following a user-centered approach, we developed the system by involving domain experts and final users starting from the first design phases. After a requirement analysis phase, we developed mock-ups which we discussed and redesigned after several focus group sessions with experts and users involved in the project. We exploited the feedback to develop our first prototype, which was first evaluated in a usability test by involving external users not cooperating at the project. After having solved the usability problems discovered in the test, the final prototype, described in this paper, was tested with real users representative of the users the Web site is devoted to; see Section 5.

As the system is devoted to the Public Administration, we had to satisfy usability and predictability requirements imposing the design of a simple user interface. Specifically, the system behavior should be highly predictable in order to support first-time visitors and to avoid frustrating professional users who would regularly use it for work purposes. Moreover, the user interface should be optimized for standard browsers.

### 3 Information about Hydric Resources

The information about hydric resources concerns rivers, lakes and underground waters and includes:

- Descriptive information about resources and observation points: e.g., maps of the points, charts representing environmental changes, pictures, documents, publications and descriptions of the monitoring stations. For instance, Figure 1 (“Caratteristiche della stazione di monitoraggio TORINO” - Features of the Torino monitoring station) provides coordinates and other information about the observation point on Po river, located in Torino, Parco Michelotti.

Consulta le informazioni chimico-biologiche su questo punto di monitoraggio

Scegli il corso d'acqua

PO

Scegli il tipo di dati che vuoi visualizzare

indici di stato ambientale\*  
valori dei parametri biologici\*  
valori dei parametri chimico-fisici e microbiologici\*

Scegli il comune

NEIVE  
PRIOLA  
ROCCA CANAVESE  
SAN MARTINO ALFIERI  
TORINO  
VENARIA

Scegli l'anno

--scegli--  
2000

\*I risultati di questa ricerca verranno generati in un file excel e in un file testuale.

**Fig. 3.** Searching qualitative data (environmental state index) about Po river in the Torino and Venaria observation points.

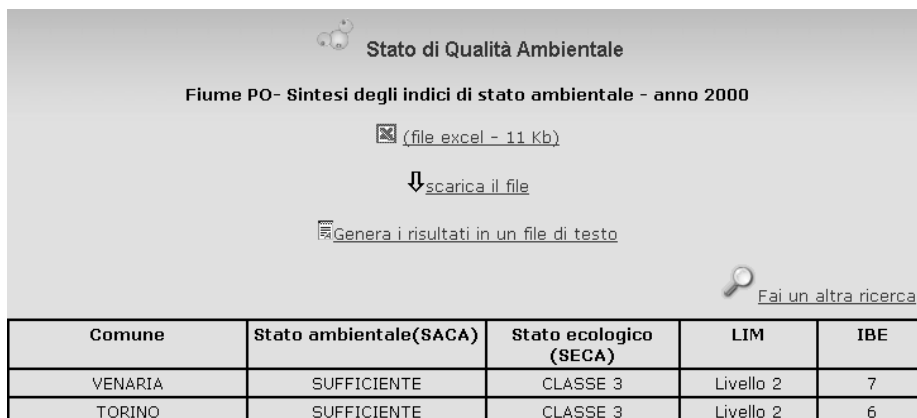
- Measurement parameters concerning physical dimensions and other features characterizing the environmental state of the resources. These parameters are further distinguished in two main classes: *Qualitative parameters* are periodically measured by technicians who visit the observation points, collect data and take samples for laboratory tests. *Quantitative parameters* are monitored by automatic stations carrying out the measurements on a daily basis.

Resources may have more than one observation point. For instance, rivers are monitored in several points to check their status in the areas they flow through. Different sets of qualitative and quantitative parameters describe the various resources. Figure 2 shows some parameters concerning rivers; for each parameter, the name, an example value and the measurement unit are reported. The data about hydric resources is stored in a relational database updated in manual or automatic way. Qualitative and quantitative parameters are separately managed to reflect the fact that they are collected in different stations, following a different temporal scheduling.

## 4 The ACQUA Prototype System

### 4.1 Basic Interaction Features

The ACQUA system (demo available at <http://acqua.di.unito.it>) supports the search for information about hydric resources, including maps, descriptive information, raw and aggregated data. We present the system by focusing on information about rivers, which is sufficiently rich to give a flavor of the overall system. The ACQUA system should serve two main categories of users. The *novice users*, such as students and generic citizens, visit the Web site on an occasional basis and are not familiar with the presented content. The *expert users*, such as technicians, farmers and the personnel of other Public Administrations, frequently visit the site and are familiar with the domain specific information. In order to take these users' interaction requirements into account, the system



**Fig. 4.** Search results: environmental state index of Po river in the Torino and Venaria points.

offers two search facilities: the *simple search* facility, not described in this paper, supports a geographical search modality and guides the user step by step in the search for information. The *advanced search* facility, targeted to the expert users, supports a faster search for information by proposing forms where the user may compose the queries in a single step. Figure 3 shows the main window of the advanced search: the user may select the river(s) to analyze (top-left menu), the type of information (top right menu) and the town(s) of the observation points (bottom-left menu). Moreover, the user can specify the temporal interval of search (year(s)).

Notice that the data archive stores partial information about hydric resources. For instance, the observation stations have been set up at different times; thus, not all the data series are complete. Moreover, critical data cannot be presented in the Web site until it is validated by human experts against measurement errors. In order to prevent the user from defining search queries that will not produce any results, the system constrains the time intervals for the search depending on the availability of data. For instance, as shown in Figure 3, the environmental state indexes describing the Po river can be searched only as far as year 2000 is concerned. The user is allowed to retrieve data at different granularity levels, e.g., inspecting on one or more observation points, or retrieving data about more or less extended temporal intervals. In addition to this type of information, the database stores some data series for which the most interesting aggregated data (e.g., minimum, maximum and mean value) are already computed and can be directly downloaded by the user.

The results of the user's queries are displayed in Web pages showing different types of information. For instance, Figure 4 ("Stato di Qualità Ambientale" - Quality of environmental state) shows the results of the query performed in Figure 3, where the user asks for the environmental state of Po river, in the Torino and Venaria observation points. The top of the page summarizes the user's query ("Fiume PO - Sintesi degli Indici ..." - Po river, synthesis of environmental state indexes - 2000). The search results are available in different formats: by following the links below the title, they can be



**Fig. 5.** Annotated search results: a link suggests follow-up queries.

downloaded as a MS Excel® table suitable for data analysis or as a plain text file (“Genera i risultati in un file di testo”). Moreover, the results are shown in a table reporting: the town of the observation points (“Comune”), the values of the environmental state (SACA), the values of the ecological state (SECA), and so forth.<sup>1</sup>

#### 4.2 Adaptive Features of the Prototype

The information provided by the ACQUA system concerns several aspects about hydric resources (see Section 3). In order to guarantee that the user can easily understand the search results, we imposed some restrictions on the queries that she can perform. For instance, as far as qualitative data about rivers is concerned, the user may request the environmental state indexes, or details about biological, microbiological and chemical parameters. However, she is not allowed to combine the queries because the results would be confusing. Therefore, if the user wants to retrieve different types of information, she has to execute more than one search query.

Although the separation of search paths enforces the clarity of the presented results, it makes the retrieval of different types of information a lengthy task. As a compromise

<sup>1</sup> Other formats, e.g., XML, could be generated to make data portable to other applications, but we did not add such formats in the current system because at the moment most users are only interested in the pictorial and in the Excel® tables.

*IF the user asks for (quantitative) historical data for more than one observation point  
THEN suggest qualitative data  
IF the user asks for chemical data for more than one observation point  
THEN suggest biological data  
IF the user asks for biological and chemical data and environmental state indexes  
THEN suggest quantitative data*

**Fig. 6.** Some association rules for the anticipation of the user’s information needs.

between clarity and efficiency, we have extended the system with a personalized search feature that supports the retrieval of information strictly related to the user's explicit search goals. When possible, the system anticipates the user's information needs and proposes supplementary *follow-up queries* as links [8] that the user can follow to retrieve the extended search results without starting the query specification from the beginning. For instance, see Figure 5: the higher portion of the figure is similar to the previous one, but the system's recommendations are proposed in the lower portion of the page ("Ti consigliamo anche i valori dei parametri chimici e microbiologici" - We also suggest results about chemical and micro-biological parameters).

The user's queries are anticipated by monitoring her browsing behavior (previous queries during the interaction) and by applying association rules that suggest other related queries. Figure 6 presents some rules in linguistic form. In the selection of the association rules to be applied, each search path followed by the user is separately treated, assuming that she may focus on different types of information during the same interaction session.

We defined the association rules (25 rules in the current system) by statistically analyzing the recurring patterns in the Natural Language queries posed by real users to the Hydric Resources Division via e-mail and fax. As the messages included free (Italian) text, we could not process them in an automated way: for each message, we manually analyzed the content and inputted the relevant information in a table supporting statistical analysis. We noticed that different kinds of information frequently occur together, depending on the user's request (e.g., qualitative and quantitative data). Therefore, we defined association rules based on the observed frequency derived from these requests. The association rules are represented as *condition - action* rules:

- The *condition* part is a boolean condition on the user's search path in the Web site. For instance, which type of data she has asked about, how many observation points she has selected for the analysis, which temporal interval she has selected, and similar.
- The *action* part is the extended query to be proposed by the system; e.g., select another type of data on the same observation points and temporal interval.

At the current stage, the association rules are mutually exclusive and are managed by employing a simple inference mechanism. In future revisions of the rule set, we might introduce priorities to handle the possible conflicts.

## 5 Evaluation of the Adaptive Features Offered by ACQUA

**Subjects.** We evaluated 10 potential users of systems, 4 females and 6 males, 30-50 aged. All the subjects work in the hydric resource field and none of them was involved in the project.

**Procedure.** The subjects were split in two groups (5 subjects each) and randomly assigned to one of the two groups. The experimental group had to solve the tasks using the adaptive Web site, while the control group had to solve tasks without adaptation.

**Experimental tasks.** Each subject had to solve 7 tasks, each one representing a real task the user can perform in the Web site. The tasks were strictly correlated and could be

grouped in three main search activities the user often performs together, as our correlation study suggests (see Section 4.2). The first activity conveyed the whole information useful to an the environmental impact study. The second one supported constructions feasibility studies. The third activity supported lawyers studies and work. In the first activity (that includes tasks 1, 2 and 3), we asked the user to search for the complete sets of qualitative data - environmental, chemical, biological - concerning two observation points on the Po river. In the second activity (tasks 4 and 5), the user was requested to search for quantitative data about the Po river in Turin, and then to search for the related information about underground water. In the third activity (tasks 6 and 7), the user had to search for quantitative data about the Po river in Turin at a specific day, and then to retrieve detailed information about the related observation point.

In the control group, the users had to submit a new query for every task, in order to obtain the requested results. In the experimental group, the users could obtain the related information by clicking on the annotated link showing the follow-up query (see Figure 5). The adaptive suggestion appeared in tasks 2, 3, 5 and 7 after an initial query generating the results to be correlated. In that way the user did not submit a new query by filling a form, but the related results were automatically generated by exploiting the association rules described in Section 4.2.

**Experimental design.** Single factor (the adaptivity) between-subjects design.

**Measures.** The subjects' performances were recorded by using Camtasia Studio®. We measured the task completion time and the subjects' satisfaction (by means of a post-task walkthrough).

**Hypothesis.** We hypothesized that the users working in the experimental group could obtain better performance results.

**Results.** The ANOVA showed that the subjects of the experimental group had better performance results. In addition, we calculated the effect size (treatment magnitude) and the power (sensitivity) as suggested in [2]. The effect size ( $\omega^2$ ) measures the strength, or the magnitude, of the treatment effects in an experiment. In behavioral sciences small, medium, large effects of  $\omega^2$  are 0,01/0,06/>0,15 respectively. The power of an experiment ( $n'$ ) is the ability to recognize treatment effects. The power can be used for estimating the sample size. In social science the accepted value of the power is equal to 0,80, which means that the 80% of repeated experiments will produce the same results.

- task 2. ANOVA:  $F(1,8) = 12,45$   $p < 0,01$ ;  $\omega^2 = 0,53$ ;  $n' = 3,49$ .
- task 3. ANOVA:  $F(1,8) = 12,12$   $p < 0,01$ ;  $\omega^2 = 0,53$ ;  $n' = 3,60$ .
- task 5. ANOVA:  $F(1,8) = 14,16$   $p < 0,01$ ;  $\omega^2 = 0,57$ ;  $n' = 3,04$ .
- task 7. ANOVA:  $F(1,8) = 9,23$   $p < 0,05$ ;  $\omega^2 = 0,45$ ;  $n' = 4,86$ .

Notice that the results are significative and we obtained a large estimate of the magnitude of the treatment effect. In addition, by exploiting a power of 0,80 and the corresponding  $\omega^2$  for each task we could determine the requested sample size, which fits our sample size ( $n=5$ ).

**Post-task walkthrough.** During a post-task walkthrough the subjects are asked to think about the event and comment to their actions. After each test we talked with the subjects in order to collect their impressions and to discuss about their performance and the problems they experienced during the text. In this session we wanted also to collect some feedback useful to produce a qualitative evaluation of the site. As a recent study

pointed out [9], statistical analyses are often false, misleading and too narrow, while insights and qualitative studies do not suffer from these problems because they strictly rely to the users' observed behavior and reactions. Even if our experimental evaluation reported significant results supporting our hypothesis, the actual behavior of the users could be different. An ethnographic study was not feasible at that time, because the information needed to feed the databases was partial. Therefore, we opted for a post-task walkthrough to collect some deeper insights.

In most cases the interviewed users were satisfied about the site. Most of them encountered some difficulties in execution of the starting query of task 2, thus we modified the interface form. All the users of the experimental group exploited the suggested adaptive link, but they were not aware of its personalized features. When we explained the adaptations, they noticed the particularity of the suggestion ("We also recommend you..."). However, they were attracted from the suggestions and they really appreciated the possibility of skipping the execution of a new query. The adaptive suggestions were considered visible but not intrusive. Also the users of the control group reported similar considerations when the adaptive features of the site were explained. Even if they did not found the suggestions in their tasks, they explored the results page in order to find some shortcuts to proceed in the task execution. After having followed some links, they went back to the previous query page or to the home page by exploiting the browser's "Back" button. In both conditions, a common behavior pattern emerged: the users explored the results page before starting a new search. Nevertheless, their behavior could be influenced by the test condition, because tested users usually pay more attention to their own actions and to the page design. Even if they were satisfied about the adaptations features, only the real usage can demonstrate our hypothesis. However, both quantitative and qualitative test results are encouraging and we think that the adaptation results are shown in the right portion of the pages. After this test, we presented the adaptive version of the site to the expert users of the Hydric Resources Division collaborating to the project. They confirmed the correctness of the association rules we exploited for the adaptation and they decided to replace the non-adaptive version of the prototype with the adaptive one.

## 6 Discussion

We presented the interactive features offered by the ACQUA prototype system to support the search of qualitative and quantitative information about Piedmont hydric resources. The system has a simple user interface, designed to meet simplicity, usability and predictability requirements, but it offers advanced interactive features enabling the user to create her own view on the information space. Two search features, targeted to novice and expert users, are available, and the search results are presented in pictorial and machine readable formats, in order to support data manipulation at the user side. Moreover, the system analyzes the user's queries to identify her information needs, and it exploits personalized association rules to propose *follow-up queries* complementing the search results with strictly related information. The follow-up queries are performed on demand; thus, the user can ignore them if she is not interested in the additional data, and the system does not need to retrieve and pre-cache uninteresting information.

The advanced search features we presented differ from the related work in various respects. On the one hand, the inferences made in our system are simpler than the probabilistic ones applied in other automated assistants, such as Lumière [5], which exploit Bayesian Networks to capture the dependencies between the user actions. The point is that the user visiting our Web site is only searching data, and she is not carrying out a complex task requiring a problem solving activity. Therefore, lightweight rules associating contextually related search queries are sufficient to predict the implicit user interests and to complement the search for information accordingly. Our approach also differs from the *follow-up question answering* techniques proposed in [8]: our follow-up queries are pre-compiled in a set of association rules, instead of being generated by a planner, in order to efficiently manage the query identification process.

On the other hand, we apply query analysis techniques to identify regularities in search patterns. This is different from the typical inferences carried out in recommender systems, which reason about regularities in the selection of individual items (e.g., see [4, 3]), or about the features of the selected items (e.g., see [1]) to identify the user's priorities. Other strategies for personalized Web search have been proposed in [6]. Different from our work, that system does not personalize the proposed results but supplies a small set of categories as context for each query. The system combines both the user's search history and a general user profile extracted automatically from a category hierarchy to offer a personalized context disambiguating the proposed query results. In our system, we did not manage long-term user preferences because we noticed that, during different interaction sections, the same users are interested in rather different types of information. We thus decided to base the system's recommendations only on the analysis of her search behavior.

Before concluding this presentation it is worth mentioning that, although the first set of adaptation rules supporting the user's search task has been manually defined, this was a lengthy work and could not be easily replicated to revise the rules along time. However, if the Hydric Resource Division adopts the ACQUA system as its official information system, the log files generated by the system will provide structured evidence about user behavior. Thus, datamining techniques could be exploited to automatically recognize usage patterns and revise the adaptation rules. In our future work, we will also exploit some ontology management techniques in order to design a domain ontology supporting the description of the user's navigation behavior at an abstraction level suitable for efficient automated analysis.

## References

1. D. Billsus and M. Pazzani. A personal news agent that talks, learns and explains. In *Proc. 3rd Int. Conf. on Autonomous Agents (Agents '99)*, pages 268–275, Seattle, WA, 1999.
2. D.N. Chin. Empirical evaluation of user models and user-adapted systems. *User Modeling and User-Adapted Interaction*, 11(1-2):181–194, 2001.
3. P. Cotter and B. Smyth. WAPing the Web: content personalization for WAP-enabled devices. In *Proc. International Conference on Adaptive Hypermedia and Adaptive Web-based Systems (AH 2000)*, pages 98–108, Trento, Italy, 2000.
4. GroupLens. GroupLens research. <http://www.ncs.umn.edu/Research/GroupLens>, 2002.

5. E. Horvitz, J. Breese, D. Heckerman, D. Hovel, and K. Rommelse. The Lumière project: Bayesian user modeling for inferring the goals and needs of software users. In *Proc. 14th Conf. on Uncertainty in Artificial Intelligence (UAI '98)*, San Francisco, CA, 1998.
6. F. Liu, C. Yu, and W. Meng. Personalized web search by mapping user query to categories. In *Proc. CIKM'02*, McLean, Virginia, USA, 2002.
7. M. Maybury and P. Brusilovsky, editors. *The adaptive Web*, volume 45. Communications of the ACM, 2002.
8. J.D. Moore and V.O. Mittal. Dynamically generated follow-up questions. *IEEE Computer*, 29(7):75–86, 1996.
9. J. Nielsen. Risks of quantitative studies. In *Alertbox*, <http://www.useit.com/alertbox/20040301.html>, 2004.