

Serving Enhanced Hypermedia Information

Abstract

An apparent limitation of existing Web pages is their inability to accommodate differences in the interests and needs of individual users. The present paper describes an approach that dynamically customises the content of public Web-based information via an interceding 'enhancement server'. The design and operation of this system is described with examples drawn from two current versions. Indications from early trials support the view that this approach affords considerable scope for accommodating the needs and interests of individual Web users.

Introduction

Perhaps the major advantage afforded to designers by the Web is the facility to create hyperlink documents and so produce an interconnected pool of information. Users commonly use links to navigate in the WWW. In a recent user survey [1], most users (17.2%) in responding to the question "How Users Find out About WWW Pages" indicated that they follow hyperlinks from other Web pages. The same survey indicates that users are often unable to find the information they seek (7.1%) and a smaller percentage (4.7%) failed to find a page that they knew existed.

This observation suggests that some difficulties that arise during users' navigation could be alleviated if the visited web pages contained hyperlinks to the information web surfers were looking for. However, this cannot always be achieved. When a web developer creates a web page, it is customary that only links relevant to the hosting site's main purpose are included. For instance, within a site providing geographic information, the word "Italy" might be linked to the prefectures or the neighbouring countries of Italy, but not to pages about its culture or economy. It is simply impractical to accommodate all user preferences and all links to related information. On the other hand if the web user had a defined set of preferences, the visited page could have been enhanced with information pertaining to the individual's interests.

In what follows, we outline a system that will enhance the content of web pages in accommodation of user preferences. Thereafter, we propose an architecture for such a

system and discuss issues related to customisation of user information and presentation. Our fourth section describes the current system development status and the last section draws conclusions and sets goals for future work.

Objectives

Shneiderman [2] points to four key phases in the operation of any integrated framework for creativity support:

- *Collect*: learn from previous works stored in digital libraries, the web, etc.
- *Relate*: consult with peers and mentors at early, middle and late stages
- *Create*: explore, compose, discover and evaluate possible solutions
- *Donate*: disseminate the results and contribute to digital libraries, the web, etc.

A web page enhancement system can support the collection and donation phases and also the creation phase. Such an enhancement system would aim to enrich web pages according to user preferences, aid in the *collection* of appropriate information, facilitate the annotation and ranking of pages by web surfers and enable the *donation* of the collated results to other users.

Enhancement Server

To accommodate these objectives we propose an enhancement server - a server that adds 'enhancements' in conjunction with 'standard' Web services. This facility will provide a locus for Web interaction that enables users to define their own preferences for content and presentation. The service will also co-ordinate user ranking of web pages and manage annotations on Web information. An additional primary role for the enhancement server will be dissemination of user-collated data to other users with similar preferences.

System Architecture

Responsible for keeping track of users' navigation and enriching web pages, the enhancement server has to intercede between the user's browser and the web server that hosts the 'public' web pages. Users will login once to the enhancement server in order to create a profile of their individual preferences. A creation form for user

profiles is shown in Figure 1. This illustrates an enhancement facility based on second language support (cf. [3]), offering local language support for a variety of linguistic features and associated presentation characteristics.

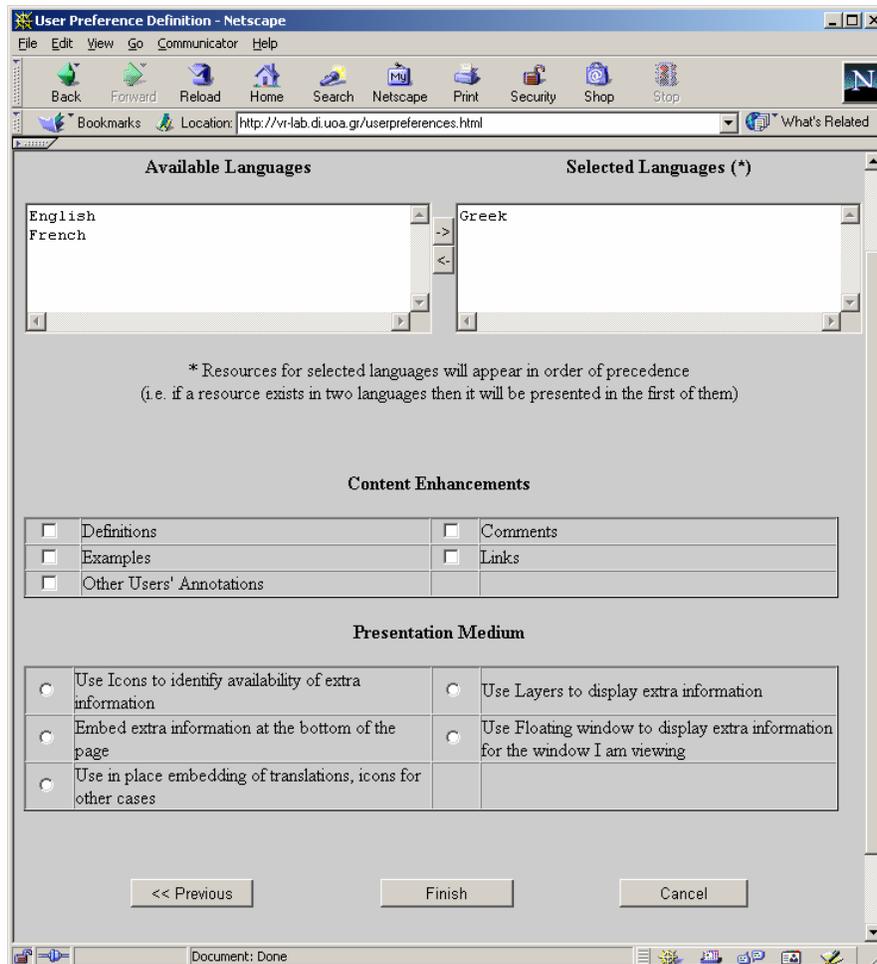


Figure 1. Setting User's Preferences

Having set their options for available enhancements (e.g., choice of local language supplements, content targets, and presentation modes) the user's browser acquires a cookie. This is the mechanism whereby returning users are subsequently identified by the enhancement server.

The enhancement server may be deployed in two user modes. In normal operation, users set their browsers to treat the enhancement server as a *proxy server*. Alternatively, users who are otherwise required to use a local proxy server (e.g., if the local administration policy requires all web traffic to pass through this proxy) may

connect directly to the enhancement server and type the URL they want to view. Operation of the enhancement server is illustrated in Figure 2.

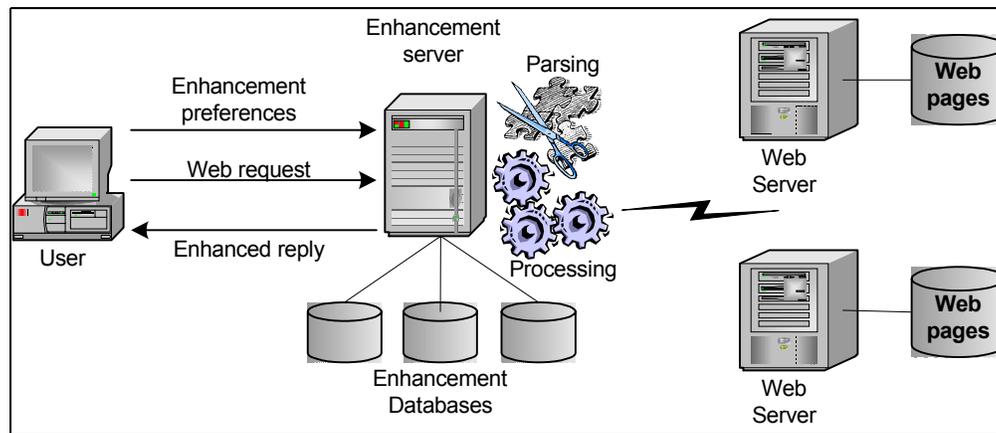


Figure 2. System Architecture

As depicted in Figure 2, the user requests a web page via the enhancement server. In response, the user receives the requested page affected by the enhancement preferences. The enhancement server scans the incoming requested web page for keywords matching user's preferences and annotates the incoming page 'on the fly' to add links to the supplemental information. On conclusion of the process, the enhanced page is sent to the user. The enhancement server comprises a parsing subsystem, a processing subsystem and several databases.

Parsing Subsystem

When the enhancement server retrieves a requested web page, the *parsing subsystem* is invoked to search for patterns that accord with the individual user's preferences. The simplest form of a pattern is a *text fragment* - either a single word or a sequence of words. If seeking to enrich pages with geographical information about countries, the text patterns *France*, *U.K.* and *United Kingdom* might be sought within the document and annotations added that link from the modified (requested) page to appropriate 'enriching' related material.

Of course, simple string searches have limited application. For instance, the European Union's legislation is commonly referred to within documents as "*EU law 02.10.10*", "*law 05.20.10.30/EU*" or simply "*09.50.10.30/EU*". In this case, the portion providing the document identity varies, and it is clearly infeasible to list all document

identities exhaustively. Such cases are accommodated through patterns that include a *regular expression* [4]. This affords a flexible means of specifying desired text portions. In this notation, the text portion corresponding to the earlier document identity can be coded as $[0-9][0-9](.[0-9][0-9])\{0,3\}$. This is interpreted as *two digits, followed by zero to three strings, each consisting of a dot and two digits*. The document identifier may be combined with constant text portions to formulate a complete regular expression that matches the desired document text, e.g. “*law [0-9][0-9](.[0-9][0-9])\{0,3\}/EU*”.

Processing Subsystem

Upon completion of the parsing phase, the results are forwarded to the processing subsystem. This operational phase inserts extra information in the page (according to user's preferences) and makes any necessary alterations for the enhancements to be displayed correctly. The processing subsystem also creates new header information and sends the enhanced Web page back to the user.

Our approach to presentation aims to leave retrieved web page looking as close to the original as possible. At the same time, the presence of extra information should be evident to the user. The presence of additional links should not interfere or obscure any original links (placed in the document by its author). To this end, the user has a number of choices regarding the presentation method for the newly embedded information:

1. An *icon* may be applied to the corresponding location in the returned document, directly **linking** to the related resource. This method works best when a small number of matches are found and there is only a single resource related to each match. Depending on the type of resource (i.e. translation, definition, example, etc) different types of icon may be used to indicate the availability of the corresponding resource.
2. An *icon* may be embedded to the corresponding location, **popping up** a new layer when the mouse hovers over this area. The new layer may host multiple items, so this method can cater for situations where multiple resources are related to each match. However, problems may arise when the document already contains layers or when the web browser does not support them.

3. Added resources may be placed at the bottom of the page (or at the top), separated from the actual page content via a horizontal line. This method is quite similar to the format used by *Google* (www.google.com) to display *cached versions* of pages or HTML versions of resources in other formats (e.g. PDF and Postscript). This technique has the advantage of supporting large volumes of added information without altering the page appearance, but, on the down side, separates the additional resources from their actual anchors within the page.
4. Extra resources may appear in a floating window. This technique preserves the appearance of the original page, but it exhibits two drawbacks. Firstly, extra information is displayed 'out of place'. Secondly, if the user has more than one browser window open, the floating window will display information for the last window opened.
5. Translations of key terms can be placed in the web page, since they stand out from the rest of the document. Other resources can use icons as described in the first approach. This alternative was used in the evaluation version of the enhancement server.

An additional task addressed by the processing subsystem is to apply the tags required for correct display of multi-lingual resources. When an additional resource in some language needs to be added to the result page, the required character set for the resource presentation is compared to the page's primary character set, designated in the `<meta http-equiv="content-type">` header tag. If these character sets are different, the inserted data are enclosed between the tags `` and ``, signifying to the browser which fonts are required to display these pieces of information.

When all supplemental information has been embedded within the retrieved page, the enhancement server computes the new document size, changes the HTTP headers accordingly and forwards the document to the user (strictly, to the user's browser).

Database Subsystem

The processing subsystem acquires the enhancement resources from the database subsystem. There may be one or more databases, depending on the thematic areas the supported by the enhancement server. The databases hold information such as key terms, definitions, comments, examples or links. The databases may also retain user

submitted annotations of concepts. Furthermore, since the information may exist in more than one language, the database subsystem can also afford translation of key terms in other languages.

Implementation and Development

Implementation of the enhancement server is based on an Apache Web Server. To date, we have realised part of the intended functionality. In the current release, users are able to set the enhancement server as their default proxy server and receive enhanced web pages. However, they are not yet able to define personal preferences. Instead, users receive a full set of content enhancements (apart from user annotations).

Two systems have been implemented to assess feasibility and usability in different application areas. The first system addresses the area of computer science - providing computer science key terms translations, definitions and examples in Greek. An illustration from this system is given in Figure 3.

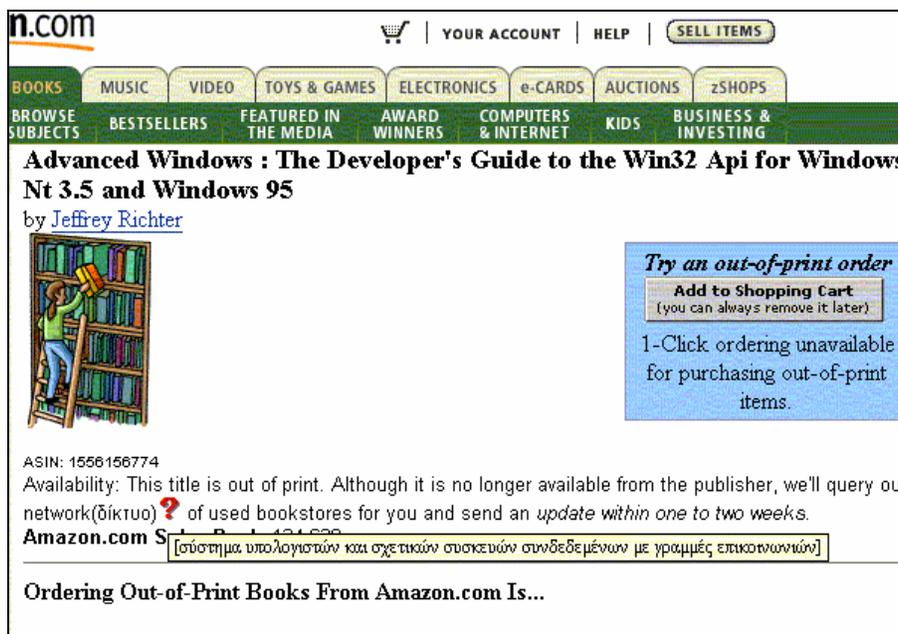


Figure 3. Web page enhanced with translation and definition of computer science terms

The second system addresses the area of European Union laws with direct links to europa.eu.int legislation. The system recognises patterns that match the format Directive xxxx/yy/EC (in any European language) and includes a link to the relevant law.

Conclusions – Future Work

In this paper we describe on-going work on a system that affords enhanced browsing of web documents. The system is currently under testing to evaluate its potential to support user interaction on the web. This spans from language specific support to subject specific extensions. First results are encouraging, and the next release of the system will allow users to personalise the support information received, both for the resource type and for the presentation style.

References

- [1] GVU's WWW Surveying Team. "10th WWW User Survey", Georgia Tech, Atlanta, GA., http://www.cc.gatech.edu/gvu/user_surveys/survey-1998-10/
- [2] Ben Shneiderman "Supporting Creativity with Advanced Information-Abundant User Interfaces", HCIL Technical Report No. 99-16 (July 1999), <http://www.cs.umd.edu/hcil>
- [3] Weir, G.R.S. and Lepouras, G., English Assistant: A Support Strategy for On-Line Second Language Learning, ICALT 2001, IEEE International Conference on Advanced Learning Technologies, Madison, USA, 2001.
- [4] MySQL group, "Description of MySQL regular expression syntax", available at http://www.mysql.com/documentation/mysql/bychapter/manual_Regexp.html