

Kent Academic Repository

Full text document (pdf)

Citation for published version

Bovey, John D. (2006) Adding User-Editing to a Catalogue of Cartoon Drawings. In: Gonzalo, Julio and Thanos, Constantino and Verdejo, M.Felisa and Carrasco, Rafael.C., eds. Research and Advanced Technology for Digital Libraries. Lecture Notes in Computer Science, 4172 . Springer, Berlin, pp. 457-460. ISBN 3-540-44636-2.

DOI

https://doi.org/10.1007/11863878_42

Link to record in KAR

<https://kar.kent.ac.uk/14412/>

Document Version

UNSPECIFIED

Copyright & reuse

Content in the Kent Academic Repository is made available for research purposes. Unless otherwise stated all content is protected by copyright and in the absence of an open licence (eg Creative Commons), permissions for further reuse of content should be sought from the publisher, author or other copyright holder.

Versions of research

The version in the Kent Academic Repository may differ from the final published version.

Users are advised to check <http://kar.kent.ac.uk> for the status of the paper. **Users should always cite the published version of record.**

Enquiries

For any further enquiries regarding the licence status of this document, please contact:

researchsupport@kent.ac.uk

If you believe this document infringes copyright then please contact the KAR admin team with the take-down information provided at <http://kar.kent.ac.uk/contact.html>

Adding User-Editing to a Catalogue of Cartoon Drawings

John Bovey

Computing Laboratory, The University, Canterbury, Kent CT2 7NF, UK
j.d.bovey@kent.ac.uk

Abstract. This paper describes an ongoing project to enable user-editing on an existing online database of about 120,000 British newspaper cartoons at the University of Kent. It describes the cartoon catalogue itself and then describes how the online search website has been extended to allow users to edit catalogue records in a way that should be both safe and economical. Finally, it discusses the next stage of the project, which is to experiment with ways to encourage users to become contributors.

1 Introduction

Online databases are usually one-way in that their data is provided and updated centrally, and is consumed by users. This is in contrast to shared resources such as, for example, the Wikipedia [2],[3],[4], where the information it contains is entered and maintained by users. It must be true, though, that for many one-way information resources, the users have a lot of information that could potentially be added to the resource and would improve it for everyone if that could be done. This user information could be simple corrections or more substantial additions. If we want to capture this information then we need to find a way to encourage users to contribute their knowledge while protecting the existing data from damage. This is what we are trying to do with our own online database of newspaper cartoons. . The system can be seen at opal.kent.ac.uk/cartoonx-cgi/ccp.py.

2 The Cartoon Catalogue

The University of Kent Centre for the Study of Cartoons and Caricature (more usually just called the Cartoon Centre) was established in 1973 when the university became the custodian of a collection of original artwork of 20th Century British newspaper cartoons. The collection, which included large bodies of work by, among others, David Low, Sidney George Strube and Victor Weisz (Vicky), was recognized as being a useful historical research resource and so the Cartoon Centre was set up and work started on cataloguing. The original catalogue was based on cards and photographs, but eventually it was replaced by an online catalogue with digitized

images. From 1998, the Cartoon Catalogue has been freely accessible as a web site [1]. The presence of digitized images makes the catalogue into a self-contained research resource.

Each cartoon has a catalogue record and one or more digital images. The records are subdivided into fields, including mechanically derivable ones like the date and place of publication, the name of the artist, the caption and any text that can be transcribed from the drawing (from speech-bubbles, for example) as well as manually added subject terms and the names of depicted personalities.

There are strong reasons why the cartoon catalogue is a good candidate for user input. One is that they are accessible public objects. Most of them appeared in newspapers or magazines and commented on events that were public knowledge at the time. No esoteric knowledge or special skills are needed to put a cartoon into context, just information that would have been widely available when the cartoon was published. Another feature of cartoons that makes them promising is that many are relevant to subjects that have their own interest groups. Cartoons on historical events (wars or elections, for example) could have information added by experts on the period when they occurred. Cartoons about sport could be put in contexts by knowledgeable sports enthusiasts, and so on. Cartoons are used as background by school teachers and could have their cataloguing improved as part of that process.

3 How User Contribution Works

A key feature of this project is that user edits are moderated by a member of the Cartoon Centre staff, but that this moderating should not be labor-intensive for staff or intrusive for contributors. The catalogue has a conventional search interface which allows the user to type in search statements and displays pages of summaries that can, in turn, be selected to see full catalogue records. A user who wants to add information to a record or correct an error clicks on the [Edit this record](#) link in the top right hand corner of the record. The record is then redisplayed as a web form with the meta-data in editable boxes. Any of these fields can be edited or added-to, and the modified record saved back to the catalogue. Changes that have been made to a record but not yet moderated are visible to the user who made them so that they can be checked, corrected or extended. Other users, however, cannot see the changes until they have been accepted by the moderator.

The moderator's main web page contains a list showing a summary of outstanding edits that are waiting to be approved or rejected. Each line in the list shows which cartoon has been edited, the fields involved, the number of words changed, and also has information about the editing history of the contributor who made the edits. In particular, it shows the number of edits they have had rejected and accepted, and a credibility score for the contributor based on their past editing record. This list can be sorted by any of these fields by clicking on the head of the column. This means that a moderator who wants to deal with the small edits first can bring them to the top of the list. Similarly, the edits by high-credibility contributors can be brought to the top of the list and dealt with quickly.

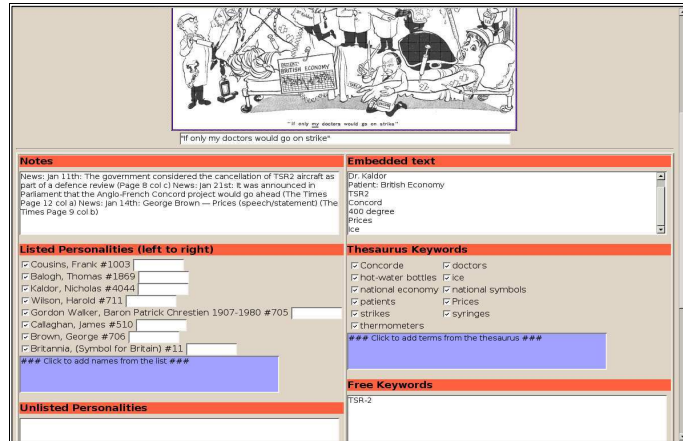


Fig. 1. The cartoon record displayed for editing

The page that the moderator uses to check individual cartoons is shown in Fig 2. The changed words are highlighted, with deletions in red and additions in green. At the top are three links: one to accept the edits as they are, one to accept them with minor changes by the moderator, and one to reject the edits outright. Clicking on the reject link puts up a web form that can be used to send an email to the contributor explaining why their edits were rejected.

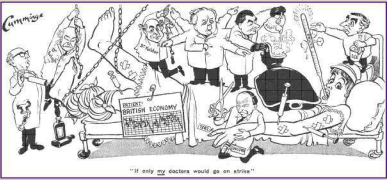
4 Implementation

The original Cartoon Centre catalogue software was developed in-house because there were no appropriate products available in 1989. This was also the case when we first put the catalogue on the web and is true of the system described in this paper. The system runs on a Linux PC and uses the Apache web server. The retrieval and editing software is written in Python, with just the time-critical parts written in C. We also use a MySQL database to store the thesaurus, name-list, contributor details and to keep track of records that have been edited but not yet moderated. The retrieval software uses inverted files and a B-tree index to provide fast searching of Boolean queries.

The catalogue records are stored in individual XML files. Each record can have one or more edit elements and one or more history elements. Each edit element represents an editing session by a contributor and contains pending changes. It is also tagged with the identification number of the contributor and the date and time when the edits were submitted. The history elements are similar, but they represent edits that have been accepted and incorporated in the main record. This means that each record contains a history of all the changes that have been made to it, when and by whom. This information can be used to generate credits to contributors or, if something goes wrong, to unwind all the edits by an individual contributor.

Record 06597 edited by Jack Brown from Canterbury, UK **ACCEPT** **ACCEPT+EDIT** **REJECT**
 Credibility 98% (8 accepted and 1 rejected)
 Edited fields: text, notes

Michael Cummings
 First published by Daily Express on 13 January 1965
 One Image
 06597 newspaper



If only my doctors would go on strike

Notes	Embedded text
News: Jan 11th: The government considered the cancellation of TSR2 aircraft as part of a defence review (Page 8 col c) News: Jan 21st: It was announced in Parliament that the Anglo-French ConcordConcorde project would go ahead (The Times Page 12 col a) News: Jan 14th: George Brown — Prices (speech/statement) (The Times Page 9 col b)	Dr. Balogh / Dr. Kaldor / Patient: British Economy / TSR2 / ConcordConcorde / 400 degree / Prices / Ice
Listed Personalities (left to right)	Thesaurus Keywords
<ul style="list-style-type: none"> Cousins, Frank #1003 Balogh, Thomas #1869 	hot-water bottles thermometers national symbols national economy doctors patients syringes ice strikes Concorde

Fig. 2. The edited record as seen by the moderator

5 Current status and Future Work

The Cartoon Catalogue has been running with user-editing for about two months and any user who takes a few minutes to register can become a contributor and edit the catalogue records, with any edits being quickly checked and incorporated into the main catalogue. Each record includes a full editing history that allows changes to be easily reversed or, more likely, acknowledged. The next stage of the project is to find ways to encourage and motivate potential editors. When we started the project, we did not know whether the major problem would be controlling over-enthusiastic contributors, protecting the catalogue against malicious edits, or encouraging people to contribute at all. It is now becoming clear that the main problem, at least initially, is encouraging people to contribute. This may be quite a challenge but we do at least now have a good test-bed for trying out different ways to encourage consumers of online data to become contributors.

References

1. Bovey, J.D., Providing Web access to a catalogue of British newspaper cartoons. Program. 37 (2003) 16-24
2. Leuf, B and Cunningham, W., The Wiki Way: Addison Wesley (2001)
3. Viegas, F.B., Wattenberg, M., Dave, K., Studying Cooperation and Conflict between Authors with History Flow Visualisations: Computer Human Interfaces Conference, Vienna, 2004 (available from web.media.mit.edu/~fviegas/papers/history_flow.pdf)
4. Wikipedia, the free encyclopedia: Available at en.wikipedia.org/wiki/Main_Page.