

# Tools for Personalised Annotation and Organisation of Diverse Web Resources

Annika Wolff

Knowledge Media Institute, The  
Open University, Milton Keynes  
A.l.wolff@open.ac.uk

Zdenek Zdrahal

Knowledge Media Institute, The  
Open University, Milton Keynes  
Z.zdrahal@open.ac.uk

Zdenek Kouba

Czech Technical University  
Department of Cybernetics  
Kouba@labe.felk.cvut.cz

## Abstract

*The CIPHER project is concerned with developing tools and methodologies to support the exploration of cultural heritage via web-based Cultural Heritage Forums. As part of the project, a Resource Annotation Tool (RAT) is being developed to support the collection, annotation and retrieval of diverse web-resources to provide content for personal or shared spaces within a forum. Interaction with the RAT tool will enable a user to reuse and personalise an official site ontology, based on personal interests and for a group of users to identify overlapping areas of interest. Tools will further be provided to facilitate users in the construction of coherent stories using their personal content, based on narrative principles. This paper describes the existing functionality of the RAT tool and plans for future development.*

## 1. Introduction

The Internet provides access to vast amounts of diverse information in different forms, such as text, images, audio and video. Whilst powerful search-engines are available to assist people in locating items of interest, there is increasing demand for portals, which provide a structured presentation of relevant information for a community of users. In particular, recent efforts have focused on producing portals for the presentation and promotion of cultural heritage. In order for such a portal to be beneficial to the target community and to fully encourage the involvement of users, it is advantageous to facilitate both personalisation of the web-space and the active participation of users in both the development and long-term maintenance of site-content. Failure to meet these requirements may lead to a reduction in usage over time.

The overall aim of the CIPHER project is to produce methodologies and tools to support the development of sustainable Cultural Heritage Forums. These forums are online spaces, accessed via a portal, where communities can actively engage with and learn about a range of heritage resources around a certain cultural theme. One particular aim is to enable forum users to develop their

own personal and shared spaces within the forum which reflect particular areas of interest and to provide tools so that the content of such spaces can be organised in a coherent manner, according to narrative principles.

Take, as an example, a virtual museum. In the physical counterpart, museum artefacts are displayed in a coherent fashion so as to tell a story. Whilst visitors to the museum are essentially free to explore artefacts in any order they like, in practice they are guided to view items in a certain order. This occurs either by explicit directives leading people along a certain route or else implicitly, due to the physical layout of the museum. The narrative is determined by the curator and it may not be apparent, or even relevant, to a visitor who does not know the deeper meaning behind the chosen arrangement of an exhibition. However, it is entirely possible that a visitor, or a visiting group, is interested in a different story to that which the curator has chosen to tell and would therefore prefer to choose their own path through the museum.

A similar situation can exist within a virtual museum, such that visitors are directed along certain pre-defined paths as they navigate through the site. However, the fact that constraints imposed by the physical layout of a museum have been removed can potentially be exploited to permit visitors to the site to create their own personalised web-space to reflect their interests, or the interests of a group. In a wider context, such as that of a Cultural Heritage Forum, it is not even necessary for artefacts to exist at one location in order for them to be included as part of a narrative. Relevant resources may exist in any location on the Internet.

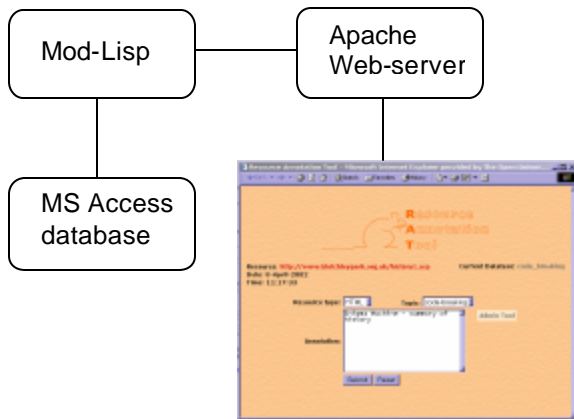
The first step in constructing a personalised web-space involves identifying, collecting and sorting the resources of interest to provide the content for the web-space. The Resource Annotation Tool (RAT), which is currently under development as part of the CIPHER project, is intended to support a user in the annotation of diverse web resources. Several different parameters can be used to describe a given resource, which is then stored for subsequent retrieval and classification. Research has shown that annotation of documents can serve several purposes in a learning context, from improving comprehension and recall, facilitating critical thinking to

recording a reader's immediate reaction to the text [1]. Further to this, the act of annotating and retrieving resources can be used to identify them as being of interest to a user and can therefore form part of an individual or group user profile.

In addition to describing the existing functionality of the tool and how it can be applied to the task of collection and retrieval of resources, the plans for future development will also be explained.

## 2. RAT architecture

The RAT tool is has been developed using Mod-Lisp, which provides an interface between the LISP programming language and an Apache web-server. This enables web pages to be dynamically generated by LISP code. There is a further SQL-based interface between LISP and an MS Access database. An overview of the RAT architecture is shown in *Figure 1*.

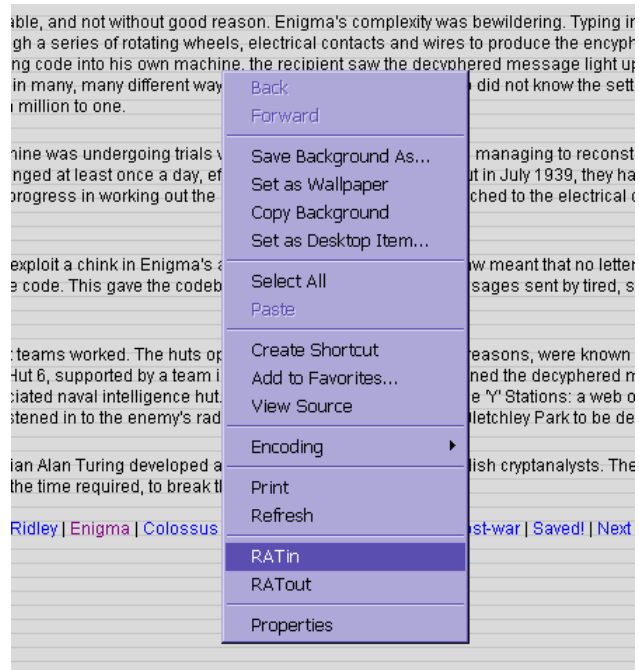


**Figure 1. The RAT tool architecture**

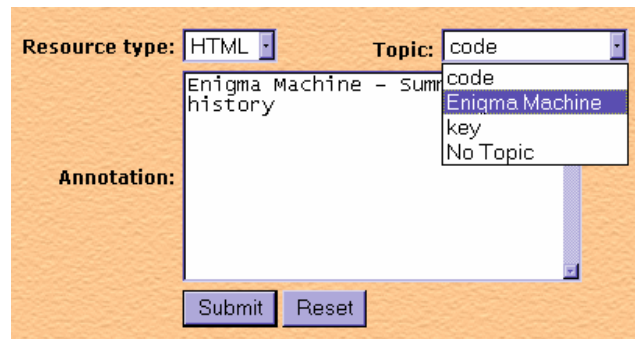
Functionally, the RAT tool is divided into two parts: 'RAT in' - for annotating and storing resources and 'RAT out' for subsequent retrieval. RAT pages are available by right-clicking the mouse button within an Internet Explorer window (*Figure 2*).

## 3. Collecting resources

The 'RAT in' tool can be used to collect and annotate any resources which are of interest to a user. At present, resources can be described by *topic*, which is selected from a menu and by *annotation*, which is entered as free-form text (see *Figure 3*). Users can define new topics using the RAT administration tool. It is also possible to specify a default topic which is then automatically selected within the topic menu. The topic, annotation and the URL of the web page are stored in an MS Access database table.



**Figure 2. Opening the RAT tool**



**Figure 3. Resource annotation - detail of Figure 1**

Future development will allow resources to be described in three ways: by *topic*, some *key-words* and a longer, *free-text annotation*. In this way, each resource will be described at different levels, ranging from the formal (topic) to the completely informal (free-text annotation).

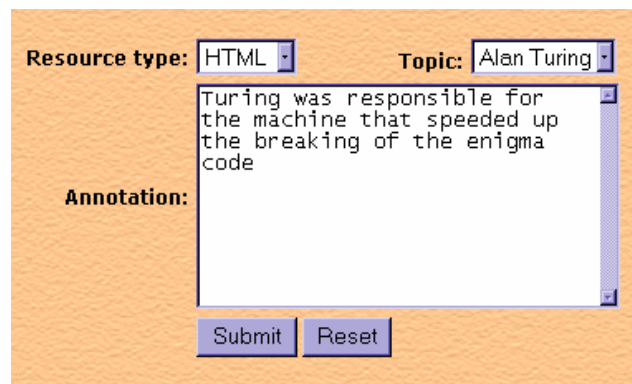
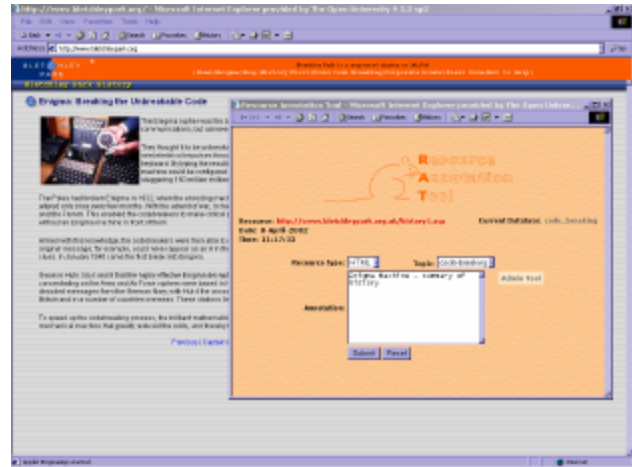
Within the context of a Cultural Heritage Forum, a topic can refer to a concept in a pre-defined ontology. An ontology is a representation of important domain specific concepts and the relations between them. Classification of resources in terms of the ontology can occur both through the specification of a topic and on the basis of the key-

words entered. Furthermore, when key-words are being used to associate resources with an ontology, it is also intended that it will be possible to make the association by finding synonyms for the key words.

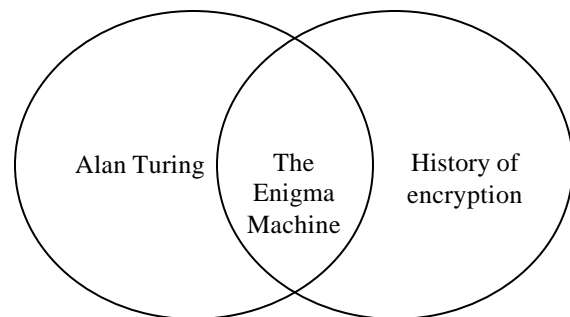
When building ontologies it is often beneficial to enable concepts and relations within the ontology to evolve and undergo a process of refinement according to feedback. Therefore, whilst the ontology can be inherited in the first instance from the site itself it can then also be updated and personalised according to user-interests. To enable this to occur, it will also be possible to define isolated topics that are not associated with the ontology, which could either be incorporated into it at a later date, left as isolated topics or else deleted. A further benefit of this approach is that it will provide users with flexibility for both storage and retrieval.

Once a user has accumulated a number of diverse resources, it is anticipated that it will be possible to apply machine-learning algorithms to the dataset in order to extract common principles from the data and to suggest new concepts for the ontology. New concepts can be integrated into personalised ontologies and/or suggested to the administrator of the official site in order to increase the scope of the forum. In this way, both personalised and official ontologies will evolve over time. It will also be possible for a group of users to apply these principles to their collective datasets in order to identify areas of common interest and to provide content for shared spaces.

As an example, *Figure 4* shows two screenshots. The first shows the annotation of a page describing the history of the Enigma Machine (the annotation is the same as that shown in *Figure 3*). The second screen shot depicts a possible annotation of the same page made by a different visitor. This illustrates how it is possible that one visitor to this site may store this page due to an interest in the history of encryption devices and code-breaking techniques and a different visitor may store the page as part of an investigation into the life of Alan Turing. It can be imagined that, aside from this particular page, there may be little overall overlap between a set of resources collected by each person. If, however, both visitors did begin to look into the Enigma machine itself in more detail, application of the learning algorithm across their datasets could highlight this common interest in the Enigma machine, and the users could collaborate in the development of a shared-space dedicated to this subject (*Figure 5*).



**Figure 4. Collecting resources about the Enigma Machine**



**Figure 5. Overlap of interests within two personal spaces**

#### 4. Ontology definition

Ontologies can be defined using the Apollo knowledge modelling tool, developed at KMi. *Figure 6* shows a partial screenshot of the Apollo tool with an example

ontology that describes encryption devices. It includes concepts such as *code*, *key* and *encryption/code-breaking-machine* and inherits properties from the two super-ontologies: *mechanical-devices* (which includes the concept of *machine*) and *time-and-date*.

This could possibly represent part of an official ontology specifically regarding the Enigma machine. As an example of how an ontology can be extended according to user-preference, in the case that a user was interested in various other encryption devices, it would be possible to extend this ontology to include other instances of encryption and decoding devices. New classes and subclasses can also be added. For instance, during the course of researching various encryption devices the user may become interested in the difference between analogue and digital devices. If this was reflected in the user's annotation of various resources, learning algorithms applied to the dataset could use this information to suggest these as new concepts for the ontology. An alternative to re-using an existing site-ontology is for the user to define it themselves. Apollo is a user-friendly tool which simplifies the input and editing of ontologies.

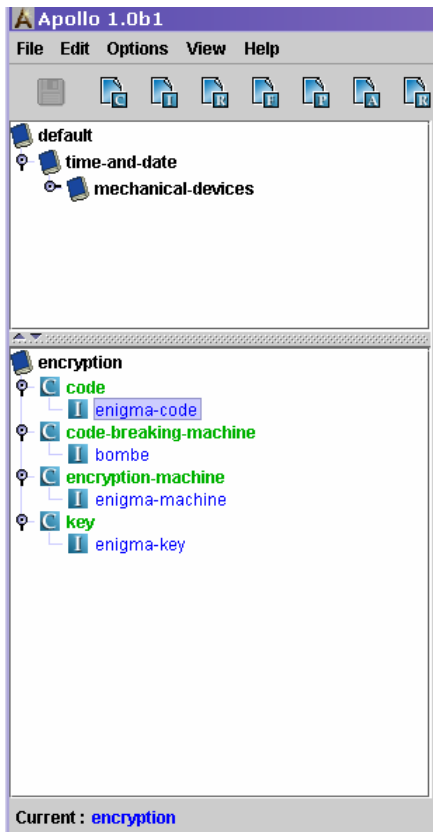


Figure 6. Encryption ontology within Apollo

## 5. Retrieving resources

Retrieval of resources is achieved using the "RAT out" tool (Figure 7). A user can specify the following parameters to search by: *words*, *resource type* and *topic*.

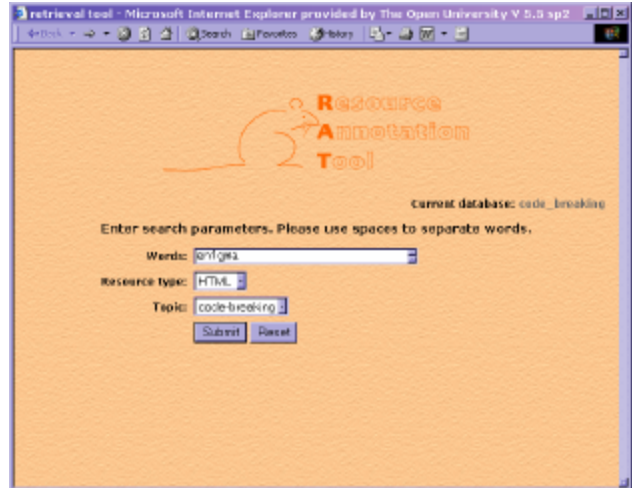


Figure 7. The 'RAT out' tool

The results are displayed as a list of matching resources (Figure 8). The free-text annotation is also displayed and can be easily modified or deleted by clicking on the "Edit" button. Clicking on the link in the URL column of the table opens the link in a new window.

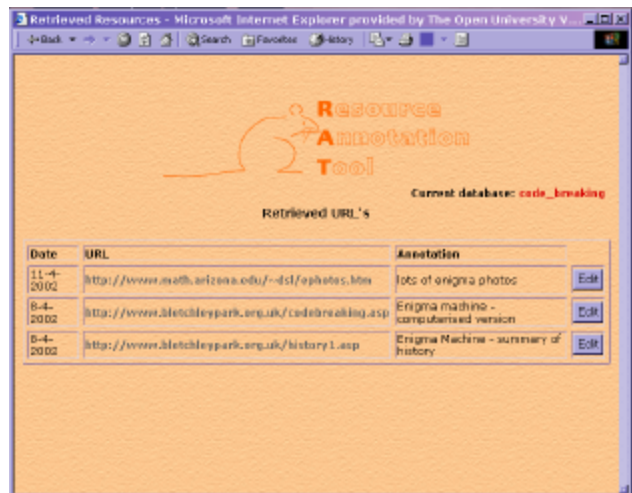


Figure 8. Retrieved resources

The tool will be further developed so that only keywords will be displayed in the resource list, with an option

provided for the user to view and edit the longer annotation. At present, the list of retrieved resources is sorted by date, with the most current resources at the top. Of course, in the future retrieval will be enhanced when resources are classified in terms of an ontology. A further refining of retrieval will occur through providing a facility whereby a user can verify the effectiveness of a retrieval strategy. "Approved" retrievals will be stored in a database containing the query and a link to the retrieved resources. Learning algorithms can then be applied to extract good clusters that can form part of a user profile and to potentially suggest new concepts for the ontology. Again, this can equally pertain to the development of shared-spaces by applying algorithms over a group dataset.

## 6. Structuring the content

It is envisaged that the RAT tool will be used to provide the content for personal and shared spaces within a Cultural Heritage forum. Further to this, it is intended that both individual users and groups should have the ability to author their own personal narratives, or to adapt existing ones, using their chosen resources. Users will be guided in the structuring of the content using narrative principles which will be defined in a formal ontology. This will enable users to provide a structure to their resources to produce coherent stories, reflecting their personal perspective of a subject. User-authored narratives can also be made available to other forum visitors.

## 7. Conclusion

The Resource Annotation Tool (RAT) is currently under development as part of the CIPHER project. It enables the collection and sorting of diverse web-resources to provide content for personal and shared spaces within a Cultural Heritage Forum. At present it supports the annotation of html, picture, audio and video resources and subsequent retrieval by key-words and topics.

Plans for future development are to associate resources with ontologies, thereby enhancing retrieval strategies and to allow refinement of ontologies by the application of learning algorithms to individual and group datasets. This can be used for two purposes, both to suggest new concepts and to highlight mutual areas of interest. The Apollo Knowledge Modelling Toolkit will be used for the input and editing of ontologies. By enabling a user to re-use and personalise an existing site ontology, pre-defined concepts will be combined with personal experience to reflect a user's interests.

The RAT tool is intended to provide only the content, rather than the presentation of resources within a personal

or shared space. However, in order to ensure that the content is coherent, it is intended that users will be guided in the organisation of resources into coherent narrative structures.

The RAT tool is currently available at <http://rat.open.ac.uk>. For further enquiries about the RAT tool please do not hesitate to contact us.

## 8. References

- [1] Wolfe, J.L. (2000). Effects of annotations on student readers and writers. *ACM Digital Libraries*, San Antonio, TX, USA.