Using a Web 2.0-style mashup approach, reusing technologies and services freely available on the web, we have developed a dynamic link service system that uses Wikipedia as its linkbase.

In our work on integrating heterogeneous cultural heritage multimedia collections, we discovered that the majority of the valuable metadata is stored in an unstructured form such as captions.

We use GATE, the General Architecture for Text Engineering, for performing the information extraction. For example:

Hungarian-born nuclear physicist Edward Teller (1908 - 2003) sits at his desk as director of the University of California’s Lawrence Radiation Laboratory in Livermore, California, 1950s. (Photo by Pictorial Parade/Getty Images)

This raises challenges in integrating the different collections, for example if people use ambiguous naming schemes, and limits the effectiveness of searching and browsing facilities.

We then do some further checks on the result to ensure we have the correct article.

Categories indicate if the article describes a person.

Article redirects provide variations on the spelling of a person’s name. For example, Wilhelm II = Kaiser Wilhelm II etc.

The text caption is submitted to the system which returns the text with a dynamically inserted link to the Wikipedia article on Edward Teller.

This enables AJAX-style link injection: when the user clicks on the "Identify People" icon, the link injection process is started.

Using Greasemonkey, a Firefox plugin that makes on-the-fly changes to specific web pages, the system can be applied to any web site.

This allows users to invoke the service as they are browsing a picture archive web site, such as Getty Images, without modifying the existing website infrastructure.

For more information, visit: http://multimedia.ecs.soton.ac.uk/semanticintegration/wikipedia-identifier/